

**UMA FAMÍLIA DE MODELOS
DE RESPOSTA AO ÍTEM
NORMAL ASSIMÉTRICA**

Jorge Luis Bazán Guzmán

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU DE DOUTOR
EM
ESTATÍSTICA

Área de Concentração: Estatística
Orientador: Prof. Dr. Heleno Bolfarine
Co-Orientador: Prof. Dr. Marcia D. Branco

Durante a elaboração deste trabalho o autor recebeu apoio financeiro da CAPES

UMA FAMÍLIA DE MODELOS DE TEORIA DE RESPOSTA AO ÍTEM
NORMAL ASSIMÉTRICA

Este exemplar corresponde à redação
final da dissertação devidamente corrigida e
defendida por Jorge Luis Bazán Guzmán
e aprovada pela comissão julgadora.

São Paulo, 17 de Junho de 2005.

Banca examinadora:

- Prof. Dr. Heleno Bolfarine (Orientador) - IME-USP
- Prof. Dr. Dalton Francisco de Andrade - IME-UFSC
- Prof. Dr. Josemar Rodrigues - UFSCAR
- Prof. Dr. Hélio dos Santos Migon - UFRJ
- Prof. Dr. Ruben Klein - LNCC/CNPQ

Non nova, sed nove
Eslogan do CÉFIRO

*Cuando comenzó a sonar
ni siquiera lo escucharon
y el leve eco
sonó a lo lejos aún más leve
Cuando al fin lo escucharon
y el lúgubre sonar de sus
cantares
a lo lejos, a lo lejos
Cuando la poesía se hizo hombre
y el niño quedo del jardín
sintióse aún más niño
y mucho más....
Cuando inexplicablemente hombre
le enseñaron la "verdad"
Cuando la sin razón tuvo una razón
Entonces, sólo entonces
había nacido nuevamente.*

Com imensa gratidão dedico
este esforço de vida

A minha família pequena: Ana, Paloma e Ernesto, grandes
em amor.

A minha família grande: Pães, irmãs, avós, tios e primos,
pequenos na distância

A meus colegas e amigos, nos caminhos da vida

Resumo

A modelagem de dados binários é dominada pelo uso de ligações simétricas (em torno da probabilidade de sucesso $p_i = 0,5$) como o logito ou o probito. No entanto, ligações simétricas podem ser inadequadas e muitas vezes mal especificadas. Na literatura, diversas propostas de ligações assimétricas tem sido desenvolvidas, entre outras, em Prentice (1976), Aranda-ordaz (1981), Guerrero e Johnson (1982), e Haro-López *et al.*(2000). Um dos propósitos desta tese foi introduzir uma função de ligação *probito assimétrica generalizada* que apresenta como casos particulares as ligações propostas em Chen *et al.* (1999), Bazan *et al.* (2004) e a ligação probito, assim como apresentar condições para as estimativas de máxima verossimilança e para a distribuição a posteriori para os parâmetros de regressão. Além disso, desenvolvemos inferência bayesiana baseada na metodologia de cadeias de Markov por Monte Carlo (Markov chain Monte Carlo - MCMC) que é de simples implementação e comparamos o modelo proposto com ligações alternativas para um conjunto de dados previamente analisado na literatura.

Por outro lado, a suposição de normalidade para variáveis latentes e curvas características dos itens (CCI) tem sido usadas nos últimos 50 anos em muitos modelos psicométricos para a Teoria de Resposta ao Item (TRI). neste trabalho, introduzimos uma nova família de modelos assimétricos para a TRI, denominada *família TRI normal assimétrica* (TRI-NA). Esta família estende o modelo probito normal, considerando a) CCIs baseadas na ligação assimétrico generalizada apresentada e b) variáveis latentes que modelam a habilidade individual distribuídas a priori com distribuição normal assimétrica. Quatro modelos fazem parte da família TRI-NA, considerando dois ou um tipo de assimetria na especificação da CCI e na especificação da variável latente. São elas: o modelo probito assimétrico-normal assimétrico (PANA), probito normal assimétrico (PNA), probito assimétrico normal (PAN) e o probito-normal usual (PN). Assim, a família TRI-NA é uma família de modelos mais flexível para se ajustar a um conjunto de dados com respostas dicotômicas. A metodologia Bayesiana de inferência é desenvolvida usando a metodologia de dados aumentados para implementar o algoritmo MCMC. Escolha de modelos entre simétricos e assimétricos também é considerada usando os critérios DIC (critério de informação de desvio), AIC e BIC esperados e residuais latentes.

A metodologia desenvolvida para a família TRI-NA é discutida e aplicada em dois conjuntos de dados. O conjunto de dados de Tanner (1996, p. 190) é usado para um estudo da eficiência na estimação dos parâmetros do modelo usando a noção de tamanho efetivo da amostra (ESS) como definido em Kass *et al.* (1998) e o tamanho efetivo da amostra por segundo (ESS/s)

como considerado em Sahu (2002). Finalmente, a metodologia é ilustrada em um conjunto de dados de um teste de Matemática aplicado a estudantes de escolas peruanas, onde é analisada brevemente a sensibilidade do modelo PAN à escolha de prioris diferentes para os parâmetros de discriminação e dificuldade dos itens.

Sugestões para a implementação de modelos usando programas livres (como BUGS) são apresentados em diversos capítulos da tese. Também, diversas sugestões para extensões dos modelos propostos tanto para regressão binária como para a TRI são proporcionadas ao longo da tese.

Conteúdo

Agradecimentos	v
Resumo	vi
Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Introdução	1
1.2 Objetivos e Organização da Dissertação	3
2 O modelo Probit Normal na Teoria de Resposta ao Item	4
2.1 Introdução	4
2.2 O modelo de resposta ao item dicotômico	6
2.2.1 Características do Modelo	6
2.2.2 Estimação	8
2.3 Inferência Bayesiana usando MCMC para o modelo TRI probito-normal	9
2.3.1 MCMC usando o algoritmo de amostragem com rejeição adaptativa (ARS)	10
2.3.2 MCMC usando amostragem de Gibbs ao considerar um esquema de dados aumentados (DAGS)	11
2.3.3 Implementação do MCMC	13
2.4 Uma aplicação	14
2.4.1 Análise de Sensibilidade usando diferentes prioris para os parâmetros dos ítems	15
2.4.2 Inferência e interpretação do parâmetros	20
2.5 Discussão Final	23
3 Uma função de ligação probito assimétrica generalizada para resposta binária	26
3.1 Introdução	26
3.2 Ligações para resposta binária	27

3.2.1	Ligações assimétricas	27
3.2.2	Ligações probito assimétricas	28
3.3	Modelo probito assimétrico generalizado	30
3.3.1	Modelo	30
3.3.2	Verossimilhança	31
3.4	Resultados principais para prioris impróprias	32
3.4.1	Existência da posteriori e da estimativa de máxima verossimilhança na abordagem condicional	33
3.4.2	Existência da posteriori e da estimativa de máxima verossimilhança na abordagem não condicional	33
3.5	Verossimilhanças aumentadas para o modelo probito assimétrico generalizado	34
3.6	Análise Bayesiana	36
3.6.1	Especificação de prioris	36
3.6.2	Esquema de dados aumentados para o amostrador de Gibbs (DAGS)	37
3.7	Uma aplicação	38
3.7.1	Comparação de modelos	39
3.7.2	Estimação de parâmetros	40
3.8	Discussão Final	42
4	A família normal assimétrica na teoria da resposta ao item	44
4.1	Introdução	44
4.2	Modelo TRI normal assimétrico	45
4.2.1	A CCI probito assimétrica	45
4.2.2	Distância entre as probabilidades de uma CCI probito assimétrica e CCI probito	47
4.2.3	Interpretação dos parâmetros dos itens na CCI probito assimétrica	48
4.2.4	Variável latente distribuída assimetricamente	51
4.2.5	A família TRI normal assimétrica	51
4.3	Estimação por máxima verossimilhança	52
4.4	Estimação Bayesiana	54
4.4.1	Especificação de Prioris	54
4.4.2	Algoritmo de dados aumentados	55
4.4.3	Distribuições condicionais completas	56
4.4.4	Implementação do MCMC	59
4.5	Probabilidade não condicional de resposta correta sob a família TRI-NA	60
5	Aplicações da família de modelos TRI-NA	61
5.1	Introdução	61
5.1.1	A implementação da estimação MCMC da TRI-NA	61

5.1.2	Critérios de Comparação de Modelos	63
5.2	Um estudo da eficiência na estimação dos parâmetros da família TRI-NA	64
5.3	Uma aplicação para um Teste de Matemática	67
5.3.1	Os dados	67
5.3.2	Uma comparação de modelos PAN e PN	67
5.3.3	O uso da família TRI-NA	70
5.3.4	A interpretação dos parâmetros	71
5.3.5	A probabilidade de resposta correta não condicional	73
5.4	Discussão	74
6	Considerações Finais	81
6.1	Conclusões	82
6.2	Futuros Trabalhos	83
A	A distribuição normal assimétrica	86
B	Critérios de comparação de modelos no contexto Bayesiano	95
C	Relações entre os modelos probito assimétricos e probito	100
D	Programas para a implementação dos modelos propostos desenvolvidos no WinBUGS	104
D.1	Estudo de eficiência nos Modelos PN e PAN	104
D.2	Análise de sensibilidade na especificação de prioris no modelo PAN	107
	Referências Bibliográficas	111

Lista de Figuras

2.1	Média posterior dos parâmetros de discriminação (a) e dificuldade (b) para diferentes prioris no modelo probito-normal no exemplo do Teste de Matemática. . .	17
2.2	Comparação de prioris (A a F) do modelo probito-normal para o exemplo do Teste de Matemática usando resumos dos coeficiente de variação das estimativas a posteriori dos parâmetros a e b sob 14 ítems	19
2.3	Histórico e densidade da cadeia para os parâmetros a e b do item 12 ($a(12)$ e $b(12)$) e média (μ) e desvio padrão (σ) das variáveis latentes u_j	21
2.4	Diagrama de caixas (“box-plot”) para parâmetros de discriminação (a) e dificuldade (b) dos itens no modelo probito-normal	24
2.5	Histograma dos escores do teste de Matemática de 14 itens para 131 estudantes	25
3.1	Curvas de probabilidade dos modelos CDS_{pa} e BBB_{pa} considerando diferentes valores para $\lambda = -5, -2, 0, 2, 5$ (nesse ordem) para escalas diferentes para $\eta = \mathbf{x}'\boldsymbol{\beta}$	31
4.1	CCI probito assimétricos para $a = 1, b = 0$ e diferentes valores do parâmetro de assimetria $\lambda = -2, -1, 0, 1, 2$ (nesse ordem)	47
4.2	CCI probito assimétricos para $a = 1, b = 0$ e diferentes valores do parâmetro de assimetria d	48
4.3	CCI probito assimétricos para diferentes valores dos parâmetros b e a considerando assimetria	50
4.4	função de densidade da normal assimétrica $\theta = (\mu, \sigma^2, \lambda)$	52
5.1	Comparação da performance das estimativas dos parâmetros para o exemplo de Tanner usando intervalos de credibilidade de 95 % da eficiência entre os modelos considerados	66
5.2	Parâmetros de discriminação e dificuldade sob os modelos PN e PAN	76
5.3	diferenças nos modelos PN e PAN $ \bar{a} + \bar{b} $ de acordo as estimativas do parâmetro d dos itens	77
5.4	Soma de quadrados de residuais latentes (SSR) para os modelos PN, PAN e PANA	78
5.5	Histogramas da distribuição posterior para o parâmetro d	79

5.6	Comparação das diferenças (estimadas-observadas) sob os modelos PN e PAN para o parâmetro de penalidade d	80
C.1	Divergência de Kullback-Leibler entre os modelos probito e <i>BBB pa</i> para diferentes valores de λ e n	102

Lista de Tabelas

2.1	Resumo de estatísticas dos escores de 14 itens no Teste de Matemática aplicado a 131 estudantes de escolas peruanas	14
2.2	Especificação de prioris usuais para os parâmetros dos ítems $\beta_j = (a_j, b_j)$ no modelo probito-normal	16
2.3	Resumo do coeficiente de variação (c.v.) das estimativas a posteriori dos a_j e b_j e Critério de Ajuste para o modelo probito-normal considerando diferentes prioris no exemplo do Teste de Matemática	18
2.4	Resumo da inferência posterior para os parâmetros sob o modelo probito-normal no exemplo do Teste de Matemática	22
3.1	Comparação dos dados sobre mortalidade dos Insetos data	40
3.2	Estatísticas posteriores dos parâmetros de regressão para diferentes modelos para os dados de mortalidade de insetos	41
3.3	Estatística posterior do parâmetro de assimetria para os modelos probito assimétricos em dados de Mortalidade de insetos	41
3.4	Desempenho de MCMC para diferentes modelos em dados de mortalidade de insetos	42
5.1	Performance na estimação de parâmetros para os dados do exemplo de Tanner (Dp: desvio padrão)	65
5.2	Comparação dos modelos TRI PAN e PN usando DIC	68
5.3	Correlação das médias a posteriori para os parâmetros dos item sob duas tipos de prioris para os parâmetros de discriminação e dificuldade	69
5.4	Comparando os modelos PN, PAN e PANA usando diferentes criterios	70
5.5	Média posterior, desvio padrão (dp) e intervalos HPD de 90 % para os parâmetros comuns sob os modelos PN e PAN	72
5.6	Média, mediana e desvio padrão (dp) a posteriori e intervalos HPD de 90 % para os parâmetros d sob o modelo PAN	73
5.7	Comparação das proporções estimadas de resposta correta sob os modelos PN e PAN com as proporções observadas na amostra	74

Capítulo 1

Introdução

1.1 Introdução

Dados são ditos binários quando cada resposta observada assume uma de duas categorias, tais como sucesso ou falha, positivo ou negativo, correto ou errado. Dados binários são comuns nas ciências sociais e médicas, mas também em agricultura, genética e ciências do comportamento. Dados binários são usados, por exemplo, na regressão binomial, modelos multiníveis, análise de dados longitudinais, meta-análise e na teoria da resposta ao ítem.

Tradicionalmente os dados binários são modelados através de funções de ligação simétricas, tais como as ligações probito e logito. Entretanto, quando a probabilidade de uma dada resposta binária se aproxima de 0 em uma taxa diferente da que se aproxima de 1, ligações simétricas são inapropriadas (ver Chen et al., 1999). Neste trabalho, propomos uma nova família paramétrica de ligações assimétricas para dados binários. O objetivo principal do trabalho é utilizar a nova ligação proposta para o desenvolvimento de uma nova família de modelos na teoria da resposta ao ítem.

A teoria da resposta ao ítem (TRI) (Hambleton e Swaminathan, 1985) é um conjunto de métodos usados na modelagem de variáveis relativas à conduta humana que substituem a teoria dos escores de um teste ou teoria clássica dos testes (Lord e Novick, 1968). Como estas variáveis não são observáveis, elas são denominadas variáveis latentes. Suposições estatísticas na modelagem da proficiência acadêmica e de outras variáveis associadas com a conduta humana estão baseadas na suposição de normalidade para a distribuição dos escores. Vários autores têm questionado esta suposição (veja Samejima, 1997 e Micceri, 1989) por considerá-la restritiva para a modelagem da conduta humana. Micceri (1989) apresenta vários exemplos de situações em que as variáveis latentes podem ser assumidas como não normalmente distribuídas. Em uma análise de características distribucionais de 440 pesquisas de aproveitamento em grandes amostras e medidas psicométricas, Micceri (1989) encontrou que 15,2 % das distribuições tinham caudas próximas da distribuição Normal, 49,1% das distribuições tinham pelo menos uma cauda destacadamente pesada, e 18% tinha ambas caudas com pesos menores que a da distribuição normal. Entre medidas de habilidade, as porcentagens foram similares com 19,5% tendo ambas caudas com pesos ao redor da distribuição normal, 57,6% tendo ao menos uma cauda pesada, e 22,9% tendo ambas caudas menos pesadas do que as da distribuição normal. Entre as medidas

psicométricas, 13,6% tinham caudas próximas da distribuição normal, 65,6% tinham ao menos uma cauda pesada, e 20,8% tinham ambas caudas menos pesadas que a Normal.

Na teoria da resposta ao item para respostas binárias são usadas funções de ligação simétricas, cuja função inversa é denominada curvas característica dos ítems. Esta suposição têm dominado a pesquisa psicométrica nos últimos 50 anos. Para contornar o problema, Samejima (2000) propôs uma família de modelos, denominada família de expoente logístico positivo cuja curva característica do item pode ser assimétrica e inclui a curva característica do item baseada na ligação logito como um caso particular. Samejima (1997) considera que curvas características dos ítems assimétricas são mais apropriadas para o modelagem da conduta humana. Samejima (1997) também indica que pode haver restrições no uso de teorias estatísticas e métodos desenvolvidos para outros campos além da pesquisa da conduta humana, em particular, aqueles baseados na suposição de normalidade. Uma necessidade sob estas circunstâncias é afastar-se da suposição de normalidade no desenvolvimento de teorias e metodologias psicométricas.

Neste trabalho, uma nova curva característica dos ítems (CCI), é assumida considerando a função de distribuição acumulada da normal assimétrica (Azzalini, 1985, Dalla Valle, 2004). Além disso, uma nova distribuição para a variável latente é também considerada assumindo a distribuição normal assimétrica. Considerando simultaneamente ou separadamente, CCI e variáveis latentes assimétricas, uma nova família de modelos TRI pode ser formada. A nova família será denominada TRI normal assimétrica (TRI-NA) e como será visto, inclui como caso especial o modelo de ogiva normal ou modelo probito-normal. Assim, estendemos os modelos TRI simétricos usuais.

A distribuição normal assimétrica é uma importante distribuição assimétrica que inclui a distribuição normal como um caso especial, mas preserva as principais propriedades da distribuição normal. Desenvolvimentos recentes nesta área incluem Azzalini e Dalla Valle (1996), Azzalini e Capitanio (1999), Sahu, Dey e Branco (2003). Um excelente panorama de trabalhos na área pode ser visto em Genton (2004). Aplicações recentes podem ser consultadas em Ozán (2003) e Lachos (2004). Importantes propriedades da distribuição normal assimétrica dada por Azzalini (1985) e Dalla Valle (2004) e novas propriedades são dadas no Apêndice A.

Nesse sentido, o presente trabalho propõe apresentar e discutir uma nova família de funções de ligação para dados binários, denominada probito assimétrico generalizada, e a aplicação desta família na teoria da resposta ao item.

Os modelos TRI são basicamente caracterizados por dois tipos de parâmetros: **ítems**, relacionados às questões ou perguntas do teste ou prova e **habilidades** (ou proficiência), relacionados com as características dos indivíduos que estão sendo medidas.

Os métodos de estimação para os modelos TRI podem ser agrupados em três categorias: estimação frequentista (ou clássica), estimação Bayesiana com e sem uso da metodologia MCMC

(Markov chain Monte Carlo). Neste trabalho nós desenvolvemos procedimentos de estimação Bayesiana usando MCMC.

1.2 Objetivos e Organização da Dissertação

O objetivo principal desta dissertação é o de propor, apresentar e discutir com detalhes uma nova família de modelos da Teoria da Resposta ao Item, a família da TRI-NA. Ela é obtida pelo uso de uma nova curva característica do item baseada no uso da nova função de ligação proposta denominada ligação proibito assimétrico generalizada e pelo uso da distribuição normal assimétrica para a distribuição das variáveis latentes associadas à conduta humana. A metodologia de estimação bayesiana é desenvolvida para os modelos obtidos e aplicações na avaliação educacional são apresentadas para os modelos propostos.

No Capítulo 2 nós apresentamos uma revisão da literatura a respeito do modelo proibito-normal na Teoria da Resposta ao Item sob uma perspectiva bayesiana, fazendo uso de metodologia MCMC. Além disso, uma aplicação é apresentada onde é desenvolvida uma análise de sensibilidade com relação a distribuição a priori. Esses modelos são bastante utilizados e possuem ampla importância na área educacional.

No Capítulo 3, nós apresentamos a ligação proibito assimétrica generalizada dentro do contexto da regressão binária. São apresentadas diversas propriedades e uma aplicação para ilustrar suas vantagens em comparação com outros modelos, usando diversos critérios de comparação de modelos.

No Capítulo 4 apresentamos a nova família de modelos de teoria da resposta ao item, denominada família TRI-normal assimétrica. Além disso, são apresentadas diversas caracterizações desta família e é discutido em detalhes a implementação da inferência bayesiana.

No Capítulo 5 apresentamos algumas aplicações da família TRI-NA. Uma análise da sensibilidade à distribuição a priori é desenvolvido, e no contexto de um conjunto de dados é discutida a interpretação dos parâmetros para o modelo proposto.

No Capítulo 6 são apresentadas diversas conclusões obtidas deste trabalho de pesquisa e sugestões para pesquisas futuras.

São também apresentados quatro apêndices a saber: A) Distribuição normal assimétrica, B) Critérios de comparação de modelos no contexto Bayesiano, C) Relações entre o modelo proibito assimétrico e o modelo proibito e D) Programas para a implementação dos modelos propostos desenvolvidos no WinBUGS.

Capítulo 2

O modelo Probit Normal na Teoria de Resposta ao Item

2.1 Introdução

A teoria da resposta ao item considera modelos estatísticos para descrever como a probabilidade de responder corretamente a uma questão (ou item) de um determinado teste ou prova depende da habilidade do indivíduo (ou proficiência) e, também, dos parâmetros relacionados as questões que estão sendo respondidas. Muitas referências na literatura estatística podem ser associadas a tais modelos, especialmente numa abordagem freqüentista, por exemplo, Baker (1992), van der Linden e Hambleton (1997) e na lingua portuguesa, o excelente resumo de Andrade *et al.*, (2000). Várias caracterizações para tais modelos têm sido desenvolvidas nos ultimos 40 anos, especialmente com respeito ao status da habilidade do indivíduo geralmente considerada como variável latente (ver, por exemplo, Borsboom *et al.*, 2003) ou, para outros autores, como parâmetros latentes (por exemplo Baker, 1992). Na abordagem desenvolvida neste capítulo, nós consideramos a caracterização devida a Holland e Rosenbaum (1986) e Bartholomew e Knuot (1999) e a abordagem Bayesiana dada em Albert (1992). Isto é, as características dos indivíduos associadas a probabilidade de resposta correta de um item são consideradas variáveis latentes e é assumida uma distribuição de probabilidade para estas variáveis.

Estimação bayesiana não é novidade nos modelos TRI e recentemente pode-se distinguir esta estimação em dois grupos; com ou sem o uso de métodos MCMC. Os métodos MCMC são um conjunto de métodos de simulação de amostras aleatórias de uma distribuição multivariada usualmente desconhecida (ver Chen *et al.*, 2000, Patz e Junker, 1999). Tais métodos são baseados na construção de uma cadeia de Markov cuja distribuição estacionária é a distribuição multivariada de interesse. No contexto da Inferência Bayesiana, a distribuição multivariada de interesse é uma distribuição a posteriori. Assim, estatísticas da distribuição teórica de interesse (desconhecida) podem ser estimadas através das correspondentes estatísticas da amostra aleatória simulada.

Os modelos TRI mais amplamente usados são os modelos de resposta dicotômica. Os modelos logito-normal e probito-normal, descritos na próxima seção, fazem parte desses mode-

los. Pesquisas de avaliação educacional conduzidas no Brasil e outros países tradicionalmente baseiam-se no modelo logito-normal que sera explicado na seção 2.1. Isto deve-se ao fato de que este modelo encontra-se implementado em pacotes comerciais e de que um outro modelo alternativo, o modelo probito-normal (veja-se seção 2.1) necessita de procedimentos de estimação mais complexos dentro de uma abordagem inferencial freqüentista. Por outro lado, no contexto Bayesiano, o uso do modelo logito-normal é algo mais complexo do que o probito-normal (ver Rupp *et al.*, 2004) devido a facilidade de conjugação quando a distribuição normal é usada.

No Brasil, Assunção (1999) implementou a metodologia dada por Patz e Junker para estimação MCMC do modelo logito-normal.

Com o uso de técnicas MCMC, o modelo probito-normal tornou-se mais popular e várias extensões foram propostas na literatura incluindo: modelos de três parâmetros (Sahu, 2002, Glas e Meijer, 2003), modelos multidimensionais (Béguin e Glas, 2001; Jackman, 2001; Linardakis e Dellaportas, 2002), modelos multiníveis (Fox e Glas, 2001), Testlet (Wang *et al.*, 2003), modelo de análise fatorial confirmatório de itens (Segall, 2003), e modelos com erro de medida (Fox e Glas, 2003).

Uma implementação da metodologia MCMC para o modelo probito-normal usando diretamente a função de verossimilhança é dada em Spiegelhalter *et al.* (1996) considerando um método de amostragem de rejeição adaptativa *adaptive rejection sampling (ARS)* (Gilks e Wild, 1992) e por Spiegelhalter, Thomas e Best (2003) considerando o amostrador de corte ou *slice sampling* (Neal, 2003). Outra implementação da metodologia MCMC, introduzida em Albert (1992), está baseada no método do amostrador de Gibbs considerando uma aumento de dados ou *data augmented Gibbs sampling (DAGS)*, usando variáveis latentes auxiliares. Embora a estimação bayesiana para o modelo probito-normal tenha sido desenvolvida na literatura, poucos trabalhos têm discutido em detalhe o problema da especificação das prioris para os parâmetros de item. Alguns resultados nesta direção são encontrados em Albert e Ghosh (2000) e Ghosh *et al.* (2000) onde é discutido teoricamente as conseqüências do uso de distribuições a priori para os parâmetros dos ítems.

Neste capítulo nós desenvolvemos uma análise de sensibilidade do modelo probito-normal com respeito à distribuições a priori especificadas e um análise da eficiência na estimação dos parâmetros dos métodos MCMC apresentados. Na seção 2.2, o modelo TRI de resposta dicotômica é apresentado bem como as suas características e procedimentos de estimação. Na Seção 2.3 é apresentada a inferência bayesiana usando MCMC para o modelo probito-normal. Também é discutida a especificação da distribuição a priori e uma formulação adequada do modelo para estimação. Na seção 2.4 é desenvolvida um análise da eficiência dos métodos MCMC apresentados e uma análise de sensibilidade com respeito à distribuições a priori usadas na literatura. Também é apresentada a inferência e interpretação dos parâmetros para um conjunto de dados provenientes de um teste de matemática.

2.2 O modelo de resposta ao item dicotômico

2.2.1 Características do Modelo

Seja

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

onde y_{ij} são os valores de Y_{ij} correspondentes as respostas dicotômicas do indivíduo i no item j e p_{ij} é uma função dos parâmetros β_j e u_i , em que $\beta_j = (a_j, b_j)'$ são os parâmetros de item, 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 à uma variável latente U_i associada ao indivíduo i , descrevendo a habilidade medida no teste. Esta habilidade o traço latente corresponde a proficiência das pesquisas em avaliação educacional. Para maiores detalhes na interpretação dos parâmetros veja-se Andrade *et al.*, (2000).

A probabilidade (condicional) de que o indivíduo i responda corretamente ao item j , $i = 1, \dots, n$, $j = 1, \dots, k$, p_{ij} 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 , $i = 1, \dots, n$, $j = 1, \dots, k$ (ver por exemplo, Johnson e Albert, 1999). Além disso, valores negativos de a_j não são esperados e b_j e U_i assumem valores reais u_i .

Fazendo $\beta_{0j} = -b_j$ e $\beta_{1j} = a_j$ temos que $m_{ij} = \beta_{0j} + \beta_{1j}u_i$ e assim u_i é um predictor linear simples da probabilidade de acerto e o vetor de parâmetros do item corresponde aos parâmetros da regressão em que β_{0j} é um parâmetro de intercepto e β_{1j} é a inclinação. Esta última parametrização é útil quando u_i corresponde a um vetor p -dimensional (MacDonald, 1997, Jackman, 2001, e Segall, 2003). Também, se fazemos $\alpha_j = a_j$ e $b_j^* = b_j/a_j$ nós temos que $m_{ij} = \alpha_j(u_i - b_j^*)$ a qual corresponde a uma parametrização mais difundida na literatura psicometrica (ver Lord, 1980, van der Linden e Hambleton, 1997).

Uma justificativa para o uso da parametrização (a_j, b_j) acima é que resulta em cálculos mais estáveis (Cook, *et al*, 2002) e apresenta maior facilidade computacional que as parametrizações (β_{0j}, β_{1j}) e (α_j, b_j^*) , não em tanto a partir desta parametrização é fácil obter os correspondentes valores dos parâmetros das outras parametrizações. Uma discussão da história da parametrizações apresentadas é dada em Baker (1992).

A função $F(\cdot)$ em (2.2) é conhecida com a *função de resposta ao item* ou *curva característica do item* e satisfaz a propriedade de *monotonicidade latente* (função estritamente não decrescente de U_i) e F tipicamente é a mesma para todo i e j . Os modelos TRI considerados tipicamente satisfazem a propriedade de *independência condicional*, isto é, para um indivíduo i , as respostas Y_{ij} , $j = 1, \dots, k$, são condicionalmente independentes dado os valores das variáveis latentes U_i , $i = 1, \dots, n$. Além disso, também é considerada a *independência entre as respostas de diferentes*

indivíduos. Sob as suposições anteriores, a distribuição conjunta de $Y = (Y'_1, \dots, Y'_n)'$ com $Y_i = (Y_{i1}, \dots, Y_{ik})'$, 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

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

O primeiro modelo TRI foi formalmente introduzido por Lord (1952) e considera $F(\cdot) = \Phi(\cdot)$, $i = 1, \dots, n$ e $j = 1, \dots, k$ com $\Phi(\cdot)$ a função acumulada da distribuição normal padrão. O modelo é conhecido como modelo de ogiva normal. Posteriormente, Birnbaum (1968) considera $F(\cdot) = L(\cdot)$ com $L(m_{ij}) = e^{m_{ij}} / (1 + e^{m_{ij}})$ denotando a função de distribuição acumulada da distribuição logística. Este modelo é conhecido como modelo logístico com dois parâmetros. Como um caso especial deste modelo ($a_j = 1$) nós temos o modelo de Rasch (Fisher and Molenaar, 1995).

O modelo TRI apresentado em (2.1)-(2.3) envolve um total de $n + 2k$ parâmetros desconhecidos tendo como consequência ser um modelo super parametrizado. De outra maneira, para um número fixo de ítems, parâmetros dos ítems são conhecidos como parâmetros *estruturais* e as variáveis latentes são conhecidas como parâmetros *incidentais*, porque eles incrementam-se quanto o tamanho da amostra n se incrementa, e porque a análise está focalizado de maneira geral nos parâmetros dos ítems. O modelo é também não identificável, desde que ele é preservado sob uma classe especial de transformações dos parâmetros (veja-se Albert, 1992) e assim as estimativas de máxima verossimilhança podem não ser únicas. Um jeito de resolver tais dificuldades é impor restrições nos parâmetros dos ítems tal como considerado, por exemplo, em Bock e Aitkin (1981). Uma outra maneira, mais de acordo com a abordagem bayesiana, é especificar uma distribuição para a variáveis latentes, tal como, considerado por Lord e Novik (1968, Chap. 16),

$$U_i \stackrel{iid}{\sim} N(\mu, \sigma^2), i = 1, \dots, n. \quad (2.4)$$

Esta suposição estabelece que nós acreditamos que as variáveis latentes são “bem comportadas” e que as habilidades são amostras aleatórias desta distribuição (ver Albert, 1992). Embora, alguma distribuição de probabilidade possa ser especificada para μ e σ^2 , como em Tsutakawa (1984), neste capítulo nós consideramos aqui $\mu = 0$ e $\sigma^2 = 1$. Esta especificação da uma escala de medida para U_i com média 0 e variância 1 e resolve o problema de *indeterminação da escala da variável latente* (Suen, 1990) e por tanto o problema de identificação dos parâmetros dos ítems (Lord, 1980).

A escolha da escala é arbitrária e uma nova escala pode ser obtida considerando transformações lineares da escala (0 – 1), especificando outros valores de μ e σ^2 . Note-se que as probabilidades condicionais em (2.2) permanecem as mesmas se diferentes escalas são utilizadas. Este fato é explicado por Andrade *et al.* (2000) como que a variável latente é invariante à escala de medida

utilizada e por tanto não faz qualquer sentido analisar ítems a partir de dos valores de seus parâmetros a e b sem conhecer a escala na qual eles foram determinados.

Suen (1990) sinalizam que conceitualmente os valores de a_j e b_j são características inerentes aos ítems e por tanto seus valores para um item dado são constantes e independentes da amostra exata de indivíduos usados para estimar esses valores. Esta propriedade dos parâmetros dos ítems é conhecida como *invariância dos parâmetros dos ítems*. Não em tanto quando os parâmetros dos ítems são estimados de uma amostra de indivíduos -num processo conhecido como *calibração dos ítems*- a invariância dos parâmetros dos ítems pode não ser aparente, isto é, os mesmos ítems calibrados através de diferentes amostras de indivíduos podem de fato mostrar diferentes valores dos parâmetros dos ítems. Isto não é por causa da perda de invariância, e sim devido a indeterminação da escala da variável latente.

O modelo TRI considerado em (2.1)-(2.4) é denominado por Albert (1992) e Lord e Novick (1968) modelo de ogiva normal. Em Bartholomew e Knoot (1999), o modelo com ligação probito e variável latente normal é denominado probito-probitto e o modelo com ligação logito e variável latente normal como logito-probitto model. Nós consideramos como mais apropriado usar a notação de modelos probito-normal e logito-normal, respectivamente. Deixando explícito o fato que a especificação de $F(\cdot)$, define uma função de ligação $F^{-1}(\cdot)$ está associado ao termo logito ou probito e o termo normal está associado a distribuição da variável latente.

2.2.2 Estimação

Os problemas de estimação podem ser agrupados em três categorias: estimação freqüentista (ou clássica), estimação Bayesiana sem MCMC e estimação Bayesiana com MCMC. A estimação clássica tem sido dominada pelo uso da estimação de máxima verossimilhança. Usando esta abordagem, várias propostas tem sido formuladas, tais como, a verossimilhança conjunta, verossimilhança marginal e verossimilhança condicional (Andrade *et al.*, 2000, Baker, 1992). Mas, como Bock and Aitken (1981) sinalizam, estimativas de máxima verossimilhança para parâmetros estruturais não são consistentes na presença de parâmetros incidentais (Neyman e Scott, 1948). Como consequência, a abordagem mais usada é o método da verossimilhança marginal, o qual usa um algoritmo pseudo-EM com quadraturas gaussianas para aproximar as integrais que são necessárias na implementação da etapa E do algoritmo (Book e Aitkin, 1981). Este procedimento de estimação está implementado em software específicos como, por exemplo, no programa BILOG (Mislevy e Bock, 1990), o que usa uma restrição "heurística" do modelo (Du Toit, 2003). A estimação dos parâmetros de habilidade é executada numa segunda etapa com os parâmetros de itens substituídos pelas suas estimativas obtidas previamente. Limitações desta metodologia são discutidas em Patz e Junker (1999) e Sahu (2002). Para citar só algumas, temos a impossibilidade de estimar U_i no caso de escores extremos (valor de escore 0 o valor

de escore n), limitações nas condições de regularidade que justificam a teoria assintótica sob a qual a inferência para os parâmetros do modelo e ajuste do modelo estão baseados (Froelich, 2000). Em situações práticas esta estimação em duas etapas define o paradigma de avaliação como no caso do pacote BILOG, o qual incorpora suas restrições computacionais como critérios da área de avaliação. Especificamente, o fato de que a estimação seja feita em duas etapas, faz com que a avaliação seja feita também em duas etapas: *o algoritmo computacional define o esquema de avaliação e não o esquema de avaliação define o algoritmo computacional*. Isto é importante por quanto o principal é definir o sistema de avaliação que seja más apropriado para um problema e logo definir o procedimento estatístico a desenvolver. Por outro lado, ainda que a estimação de parte dos parâmetros, considerando conhecidos uma outra parte, seja possível como acontece no procedimento de Bock e Aitkin (1981), a inferência baseada em argumentos assintóticos poderia resultar numa estimação não consistente dos parâmetros do modelo (Patz e Junker, 1999), especificamente na TRI nós temos observações independentes más no identicamente distribuídas e teoria assintótica para estimadores de máxima verossimilhanca marginal ainda não esta justificada (Froelich, 2000).

No caso da estimação Bayesiana (sem MCMC), é usada a estimação bayesiana marginal para as variáveis latentes considerando-se modelos hierárquicos ou não (veja Baker, 1992; Kim *et al.*, 1994) com as estimativas baseadas na distribuição a posteriori, tais como o máximo a posteriori e o valor esperado a posteriori. Uma característica das abordagens clássicas e bayesiana sem MCMC é que elas estão baseadas tipicamente na função de ligação logito. Uma maior revisão em detalhe das abordagens clássica e bayesiana sem MCMC pode ser encontrada em Baker (1992) e Andrade *et al.* (2000). Uma apresentação geral da abordagem bayesiana usando MCMC pode ser encontrada em Patz e Junker (1999). Esta última abordagem para o caso do modelo probito-normal considerando MCMC é descrita na seção seguinte e é o foco de interesse nesta pesquisa.

2.3 Inferência Bayesiana usando MCMC para o modelo TRI probito-normal

Para a inferência bayesiana é importante a especificação da função de verossimilhança e das distribuições a priori dos parâmetros do modelo.

Seja $D_{obs} = \mathbf{y}$, os dados observados. Então, a função de verossimilhança para o modelo probito-normal é dada por

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

em que $\Phi(\cdot)$ é a função acumulada da distribuição normal padrão e $m_{ij} = a_j u_i - b_j$.

Nós consideramos a seguinte classe geral de distribuições a priori independentes para os parâmetros do modelo:

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

Além disso, seguindo propostas usualmente consideradas, assumimos independência a priori entre as habilidades dos indivíduos, e os ítems. Assim,

$$\pi(\mathbf{u}, \boldsymbol{\beta}) = \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^k g_{2j}(\beta_j),$$

em que $g_{1i}(\cdot)$, $i = 1, \dots, n$ é a densidade da $N(0, 1)$. Adicionalmente, supomos que $g_{2j}(\beta_j) = g_{21j}(a_j)g_{22j}(b_j)$, em que $g_{21j}(\cdot)$ deve ser uma distribuição própria para garantir uma distribuição a posteriori própria para a distribuição conjunta $(\mathbf{u}, \boldsymbol{\beta})$ (veja Albert e Ghosh, 2000 e Ghosh *et al.* 2000). Distribuições próprias são aquelas cuja função de densidade de probabilidade integra um. Note que a priori de β_j no depende do tipo de função de ligação usada.

Seguindo propostas usualmente consideradas (veja-se Jonhson e Albert, 1999), nós consideramos $g_{21j}(\cdot)$, $j = 1, \dots, k$ como a densidade da $N(\mu_a, \sigma_a^2)$ com μ_a positivo, e $g_{22j}(\cdot)$, $j = 1, \dots, k$, como a densidade da $N(0, \sigma_b^2)$. Em conseqüência, $g_{2j}(\cdot)$ é a densidade da $N_2(\mu_\beta, \Sigma_\beta)$, $j = 1, \dots, k$ que denota uma distribuição normal bivariada com vetor de médias $\mu_\beta = (\mu_a, 0)'$ e matriz de covariância $\Sigma_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$. Não em tanto, considerando o fato que $a_j > 0$, muitos autores (Rupp *et al.*, 2004, Sahu, 2002, Patz e Junker, 1999) consideram $g_{21j}(\cdot)$, $j = 1, \dots, k$ como a densidade da $N(\mu_a, \sigma_a^2)I(a_j > 0)$ que denota a uma distribuição normal truncada para valores positivos ou como a densidade da $LN(\mu_a, \sigma_a^2)$ que denota uma distribuição log-normal. Além disso, quando especificada a a priori para b_j^* no lugar de b_j outros autores podem usar a mesma priori para u_i considerando que estão na mesma escala.

2.3.1 MCMC usando o algoritmo de amostragem com rejeição adaptativa (ARS)

Considerando a função de verossimilhança e as prioris especificadas acima, a densidade conjunta é dada por:

$$f(\mathbf{u}, \boldsymbol{\beta} | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^k g_{2j}(\beta_j)$$

ou

$$f(\mathbf{u}, \boldsymbol{\beta} | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n u_i^2 + \frac{1}{\sigma_b^2} \sum_{j=1}^k b_j^2 + \frac{1}{\sigma_a^2} \sum_{j=1}^k (a_j - \mu_a)^2 \right) \right]$$

Desta maneira, o modelo probito-normal pode ser ajustado usando MCMC, por exemplo,

através do software WinBUGS (veja-se exemplo LSAT em Spiegelhalter *et al*, 1996). Como a densidade a posteriori conjunta acima é complexa, note que, todas as distribuições condicionais completas não são padrões e, portanto, uma implementação direta usando o amostrador de Gibbs não é possível. Mas, todas as distribuições condicionais completas para o modelo probito-normal são log-côncavas (o logaritmo de suas densidades é uma função côncava) de acordo com Sahu (2002). Portanto, uma amostragem considerando distribuições unidimensionais log-côncavas pode ser feita usando o algoritmo de amostragem por rejeição, ainda que as constantes normalizadoras sejam desconhecidas (Gilks e Wild, 1992). Esses autores também desenvolveram um algoritmo de amostragem baseado num esquema de rejeição adaptativa ou Adaptive Rejection Sampling (ARS). O ARS constrói dinamicamente dois envelopes (um inferior e outro superior) para a distribuição a ser amostrada pelo uso de avaliações sucessivas da densidade nos pontos a serem rejeitados. O algoritmo pára quando um ponto proposto é aceito, reiniciando o ciclo de busca por um novo ponto. Também, no caso do parâmetro ter um intervalo de variação especificado, outro procedimento MCMC que pode ser usado é o amostrador de corte (*slice sampling*) (Neal, 2003) o qual é um método MCMC que se adapta as características das distribuições a ser amostradas e pode ser construído usando o princípio de que nós podemos amostrar de uma distribuição por amostragem uniforme da região sob o gráfico de sua função de densidade. Assim, uma cadeia de Markov que converge para sua distribuição estacionária pode ser construída na direção vertical com amostragem uniforme no corte (slice) horizontal definido pela posição usual vertical, ou de maneira mais geral, com alguma atualização que leva a uma distribuição uniforme sobre o corte (slice) invariante. Quando o modelo probito-normal é implementado em WinBUGS, o método ARS é o procedimento padrão na versão 1.3 e o método Slice Sampling na versão 1.4.

Na próxima seção, nós faremos uma descrição detalhada de um terceiro algoritmo baseado na introdução de variáveis latentes, conhecido como DAGS e inicialmente proposto por Albert (1992).

2.3.2 MCMC usando amostragem de Gibbs ao considerar um esquema de dados aumentados (DAGS)

Proposição 1. (Albert, 1992) *O Modelo TRI probito-normal, envolvendo k itens e n indivíduos, com $y_{ij} \sim \text{Bern}(p_{ij})$ e $p_{ij} = \Phi(m_{ij})$, na qual $m_{ij} = a_j u_i - b_j$, é equivalentemente definido considerando que*

$$Z_{ij} = m_{ij} + e_{ij}, \quad (2.5)$$

$$e_{ij} \sim N(0, 1), \quad (2.6)$$

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0. \end{cases} \quad (2.7)$$

Prova. Considere $F(\cdot)$ igual a $\Phi(\cdot)$ em (2.2). O resultado segue basicamente ao notar-se que $p_{ij} = P[Y_{ij} = 1|u_i, \beta_j] = P[Z_{ij} > 0|u_i, \beta_j] = \Phi(m_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, k$, pois $Z_{ij}|u_i, \beta_j \sim N(m_{ij}, 1)$. \square

O resultado mostra que uma estrutura normal linear para a variável latente auxiliar Z_{ij} leva a um modelo equivalente ao modelo probito-normal. Usando esta nova formulação e considerando $D = (\mathbf{z}, \mathbf{y})$ como os “dados aumentados” com $\mathbf{Z} = (z_{11}, \dots, z_{nk})'$, o vetor aleatório, segue que a função de verossimilhança dos dados aumentados é dada por

$$L(\mathbf{u}, \boldsymbol{\beta}|D) = \prod_{i=1}^n \prod_{j=1}^k \phi(Z_{ij}; m_{ij}, 1) I^*(Z_{ij}, y_{ij}). \quad (2.8)$$

onde $\phi(t; \mu, \sigma^2)$ denota a densidade na distribuição normal da variável t com média μ e variância σ^2 e $I^*(Z_{ij}, y_{ij}) = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$, $i = 1 \dots, n$, $j = 1, \dots, k$, é toma valor zero ou 1 dependendo do valor da variável indicadora $I(\cdot)$ que assume valor 1 se seu argumento é verdadeiro, e valor zero em caso contrário.

Proposição 2. (Johnson e Albert, 1999 *Considerando a função de verossimilhança de dados aumentados dada em (2.8) e as distribuições a priori dadas na Seção 2.3 considerando uma priori normal para a , para o Modelo TRI probito-normal as distribuições condicionais completas, necessários para o amostrador de Gibbs, são dadas por:*

- $Z_{ij}|u_i, \beta_j, D_{obs} \propto N(m_{ij}, 1)I(Z_{ij}^*, y_{ij})$, $i = 1 \dots, n$, $j = 1, \dots, k$.
- $u_i|\mathbf{Z}_i, \boldsymbol{\beta}, D_{obs} \propto N(m_{u_i}, v_{u_i})$, $i = 1 \dots, n$
com $m_{u_i} = \frac{\sum_{j=1}^k a_j (Z_{ij} + b_j)}{\sum_{j=1}^k a_j^2 + 1}$, $v_{u_i} = \frac{1}{\sum_{j=1}^k a_j^2 + 1}$, $i = 1 \dots, n$.
- $\beta_j|\mathbf{u}, \mathbf{Z}_j, D_{obs} \propto N_2(\mathbf{m}_{\beta_j}, \mathbf{v}_{\beta_j})$, $j = 1, \dots, k$,
com $\mathbf{m}_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_{\beta}^{-1}]^{-1} [\mathbf{W}'\mathbf{Z}_j + \Sigma_{\beta}^{-1}\boldsymbol{\mu}_{\beta}]$, $\mathbf{v}_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_{\beta}^{-1}]^{-1}$, em que $\boldsymbol{\mu}_{\beta} = \begin{pmatrix} \mu_a \\ 0 \end{pmatrix}$, $\Sigma_{\beta} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}$ e $\mathbf{W} = (\mathbf{u}, -1)$ onde $\mathbf{W}'_i = (u_i, -1)$, $i = 1 \dots, n$.

Prova De (2.8) e das priors especificadas na seção 2.3.1 a distribuição posterior conjunta é dada por

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

Para obter a distribuição condicional completa para Z_{ij} é suficiente considerar os termos correspondentes, i.e

$$f(\mathbf{Z}|\mathbf{u}, \boldsymbol{\beta}, D_{obs}) \propto \exp\left[-\frac{1}{2}\left(\sum_{i=1}^n \sum_{j=1}^k (Z_{ij} - m_{ij})^2\right)\right] I(Z_{ij}, y_{ij})$$

que corresponde ao núcleo de nk distribuições normal truncada univariadas para cada Z_{ij} dependendo do sinal de Z_{ij} .

Para obter a distribuição condicional completa para \mathbf{u} é suficiente considerar os termos correspondentes, i.e

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

que corresponde ao núcleo de n distribuições normal univariadas para U_i com a média e desvio padrão dados, os quais são obtido completando quadrados convenientemente.

Para obter a distribuição condicional completa para $\boldsymbol{\beta}$ é suficiente considerar os termos correspondentes, i.e

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

que corresponde ao núcleo de k distribuições normal bivariadas para β_j . \square

Com as distribuições condicionais acima, é possível implementar um amostragem de Gibbs, considerando que foi necessário a introdução de variáveis latentes o algoritmo é denominado DAGS e pode ser facilmente implementado. Rotinas em R (veja o pacote MCMC por Martin e Quinn, 2003), Matlab (veja-se Johnson e Albert, 1999) são disponíveis na Web. No WinBUGS a implementação dos procedimentos anteriores não é direta e requer uma correta especificação para as variáveis indicadoras (Ver Apêndice D para o programa correspondente). Também, quando especificada uma distribuição a priori para a_j normal truncada o Log normal o resultado da proposição 2 é direto aproveitando a conjugação destas distribuições com a normal. Assim quando considerada uma distribuição normal truncada nós acrescentamos uma função indicadora adicional $I(a_j > 0)$ e quando considerada a distribuição log normal, a_j é substituído por $\ln a_j$ na expressão correspondente. Também o resultado é bastante geral sim outras parametrizações em m_{ij} como dadas na seção 2.2.1 são usadas.

2.3.3 Implementação do MCMC

Para desenvolver procedimentos computacionais para o modelo TRI probito-normal nós podemos escrever as seguintes estruturas hierárquicas baseados nas Seções 2.3.1 a 2.3.2. Considerando o algoritmo de amostragem ARS nós trabalhamos diretamente na verossimilahnça de Bernoulli, ou seja,

$$y_{ij}|u_i, a_j, b_j \sim \text{Bern}(p_{ij});$$

$$p_{ij} = \Phi(a_j u_i - b_j);$$

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

$$a_j \sim g_{21j}(\cdot);$$

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

Considerando o amostrador DAGS, nós temos a hierarquia

$$Z_{ij}|y_{ij}, u_i, a_j, b_j \sim N(a_j u_i - b_j, 1)I(Z_{ij}, y_{ij});$$

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

$$a_j \sim g_{21j}(\cdot);$$

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

em que $g_{21j}(\cdot)$ pode ser especificado considerando as propostas apresentadas acima, i.e considerando uma distribuição normal, normal truncada ou lognormal.

2.4 Uma aplicação

Nós ilustramos a proposta bayesiana usando MCMC para o modelo probito-normal desenvolvido neste capítulo, usando um conjunto de dados provenientes de um Teste de Matemática aplicado em escolas Peruanas. A especificação de prioris, valores iniciais para definir os estados iniciais das cadeias de Markov, e diagnóstico de convergência para as cadeias de Markov são discutidas na implementação do MCMC usando o esquema DAGS.

Nesta aplicação, 14 itens de múltipla escolha do teste de matemática foram aplicados a 131 estudantes de nível sócio econômico alto em escolas de ensino básico no Peru (UMC, 2001). Vetores de respostas com respostas corretas e incorretas para os itens são apresentados no Apêndice D. O resumo das estatísticas dos escores do teste de Matemática ($i = 1, \dots, n$, $j = 1, \dots, k$) é apresentado na Tabela 2.1.

Tabela 2.1 *Resumo de estatísticas dos escores de 14 itens no Teste de Matemática aplicado a 131 estudantes de escolas peruanas*

Estatística	valor	Estatística	valor
Média	10.840	Mediana	11
Variância	3.432	Desvio padrão	1.853
Assimetria	-0.795	Curtose	0.449
Mínimo	5	Máximo	14
Alpha	0.481	P médio	0.774

Considerando as estatísticas, os escores apresentam uma média aproximada de 11 pontos e um desvio padrão aproximado de 2 pontos, assimetria negativa com presença de escores altos e uma proporção média dos ítems de 0,774 indicando um teste fácil para o grupo de respondentes. Também considerando o valor da estatística alfa de cronbach (Lord e Novick, 1968), a qual é uma medida da consistência interna dos ítems, nós achamos que o teste apresenta um nível médio de confiabilidade ou fidegnidade.

2.4.1 Análise de Sensibilidade usando diferentes prioris para os parâmetros dos ítems

Uma parte chave de qualquer análise estatística é a verificação do modelo, ou a avaliação do ajuste do modelo. Antes de se tirar qualquer conclusão a respeito da aplicação de um modelo estatístico para um conjunto de dados, um pesquisador deveria avaliar o ajuste do modelo para se assegurar que o modelo explica adequadamente importantes características do conjunto de dados (Stern e Sinharay, 2004).

Em inferência bayesiana, um pesquisador pode “conferir” o ajuste do modelo considerando os seguintes procedimentos (Gelman *et al.*, 1996): (1) checando se a inferência posterior é razoável, dado o contexto substantivo do modelo; (2) examinando a sensibilidade da inferência à mudanças razoáveis na distribuição a priori e/ou na verossimilhança; e (3) checando se o modelo pode explicar os dados, ou em outras palavras, que o modelo seja capaz de gerar dados como os dados observados. Para conferir o ajuste do modelo probito-normal, neste capítulo, nós consideramos uma análise de sensibilidade da inferência Bayesiana usando MCMC para diferentes especificações na distribuição a priori dos parâmetros dos ítems usando o conjunto de dados do teste de Matemática dados descritos antes.

O procedimento de estimação bayesiana usando DAGS como proposto na seção 2.3.2 e 2.3.3 foram implementados em WinBUGS. Cadeias com 50.000 iterações foram geradas considerando saltos (“thin”) iguais a 1, 5 e 10, e descartando as primeiras 500 iterações como em Albert (1992). Assim, os tamanhos efetivos da amostra foram 49.500, 9.900 e 4.950, respectivamente. Quando é usado MCMC, os valores amostrais para as iterações iniciais da cadeia são desconsiderados devido a sua dependência do estado inicial e para garantir a convergência à distribuição estacionária. Além disso, presença de autocorrelação entre as variáveis latentes auxiliares é observada no procedimento DAGS (Chen *et al.* 2000), por tanto valores de salto acima de 10 são recomendados para garantir um amostra de valores independentes da distribuição a posteriori.

Como mencionado na Seção 2.3.1, prioris próprias para a_j e b_j garantem que as posteriores para o modelo sejam próprias. Albert e Ghosh (2000) argumentam que a escolha de uma distribuição a priori própria para o traço latente resolve os problemas de identificação do modelo; além disso, distribuições a priori informativas colocadas para a_j e b_j podem ser usadas para refletir uma opinião a priori de que os valores dos parâmetros do item não são extremos (na fronteira do espaço paramétrico). Nas situações comuns em que pouca informação a priori é disponível

sobre os parâmetros de dificuldade, nós podemos escolher valores grandes para σ_b^2 . Esta escolha pode ter um efeito modesto na distribuição posterior para dados não extremos, e pode resultar em distribuições a posteriori próprias quando dados não extremos são observados, onde alguns estudantes apresentam respostas corretas ou incorretas para todos os itens, Albert e Ghosh (2000) e Sahu (2002), estabelecem que a especificação de valores grandes da variância σ_b^2 levam a estimativas instáveis. Estas prioris são denominadas prioris vagas (ver, por exemplo, Roberts, 2001).

Na Tabela 2.2 são apresentadas algumas prioris consideradas na literatura para os parâmetros dos itens $\beta_j = (a_j, b_j)$ no modelo probito-normal. Nós utilizamos a notação $N(0,1)I(0,)$ para indicar uma distribuição normal truncada para valores negativos, com média 0 e variância 1.

Tabela 2.2 *Especificação de prioris usuais para os parâmetros dos itens $\beta_j = (a_j, b_j)$ no modelo probito-normal*

Priori	Autor	a_j priori	b_j prior
A	Johnson e Albert (2000)	$N(2,1)$	$N(0,1)$
B	Congdon (2001)	$N(1,1)$	$N(0,1)$
C	Albert e Ghosh (2000)	$N(0,1)$	$N(0,1)$
D	Sahu (2002); Albert e Ghosh (2000)	$N(0,1)$	$N(0,10000)$
E	Spiegelhalter <i>et al.</i> (1996)	$N(0,1)I(0,)$	$N(0,10000)$
F	Sahu (2002); Patz e Junker (1999)	$N(1,0.5)I(0,)$	$N(0,2)$

As prioris *A*, *B* e *C* são prioris mais precisas (menos vagas). As prioris *D* e *E* designadas para os parâmetros de dificuldade são prioris mais difusas (vagas). As prioris *E* e *F* designadas para os parâmetros de discriminação são prioris truncadas.

Estimativas dos parâmetros de discriminação a_j e parâmetros de dificuldade b_j baseadas na média a posteriori para prioris diferentes dadas na Tabela 2.2 são mostradas na Figura 2.1. De acordo com esta figura, o modelo probito-normal é insensível as especificações das prioris para os parâmetros de dificuldade e de discriminação.

Para comparar os modelos correspondentes as seis diferentes distribuições a priori consideradas, nós calculamos a esperança a posteriori do desvio ou *posterior mean of the deviance* ($Dbar$), o desvio médio a posteriori ou *deviance of posterior mean* ($Dhat$), o número efetivo dos parâmetros ou *the effective number of parameters* (ρ_D), e o critério de informação do desvio ou *Deviance Information Criterion* (DIC), como considerado em Spiegelhalter *et al.* (2002). Eles afirmam que o DIC , como implementado no pacote WinBUGS, pode ser usado para comparar modelos complexos e grandes diferenças no critério podem ser atribuídas a diferenças preditivas reais nos modelos. Assim, o modelo com menor DIC é estimado como o modelo que deveria prever melhor um conjunto de dados replicados da mesma estrutura dos dados observados

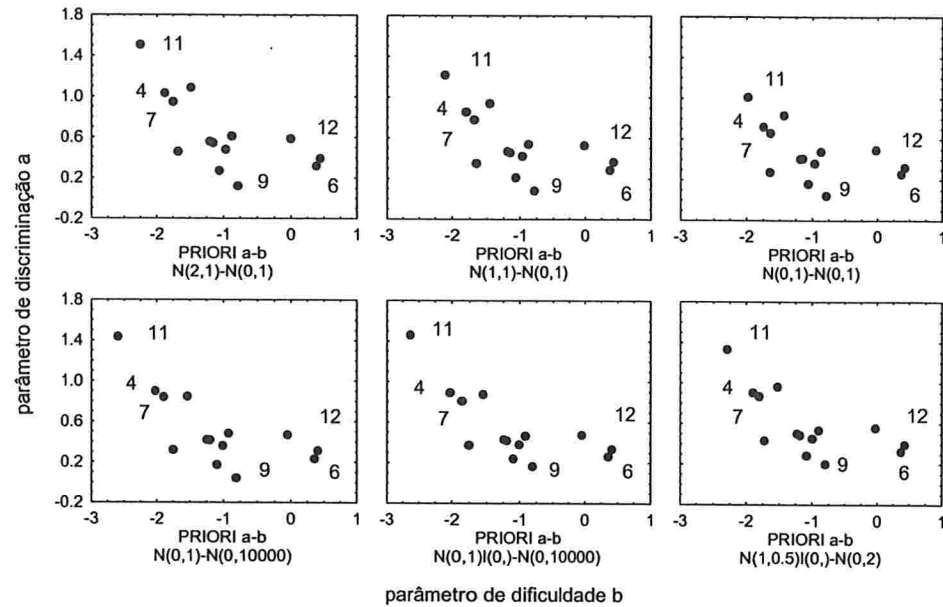


Figura 2.1 Média posterior dos parâmetros de discriminação (a) e dificuldade (b) para diferentes priors no modelo probito-normal no exemplo do Teste de Matemática.

atualmente. Esta abordagem recebeu críticas de alguns autores (veja discussão em Spiegelhalter *et al.*, 2002).

Para o cálculo do do DIC necessitamos de $Dbar$, que é o valor esperado de $D(\beta, \mathbf{u}) = -2\ln(p(\mathbf{y}|\beta, \mathbf{u}))$, i.e $Dbar = E[D(\beta, \mathbf{u})]$ a qual é uma média calculada considerando todos as estimativas a posteriori do desvio obtidas nas diferentes cadeias geradas onde \mathbf{y} inclui todos os valores dados (i.e. a amostra), e β, \mathbf{u} são as estimativas a posteriori dos parâmetros de item e das variáveis latentes nas diferentes cadeias geradas. $Dhat$ é o valor de $D(\cdot)$ acima avaliado no valor esperado a posteriori para (β, \mathbf{u}) , i.e $Dhat = D(E[(\beta, \mathbf{u})])$, ρ_D , denominado número efetivo de parâmetros é dado por $\rho_D = Dbar - Dhat$, e finalmente, DIC é dado por $DIC = Dbar + \rho_D = Dhat + 2\rho_D$. (ver Apêndice B para maiores detalhes).

Na modelagem hierárquica como acontece no modelo probito-normal a “verossimilhança” ou “complexidade do modelo” não é única, assim, o desvio também não (o mesmo valendo para DIC e ρ_D que são baseadas nesta) e pode ser calculada de vários modos (Delorio e Roberts, 2002). Quando variáveis latentes auxiliares \mathbf{Z} são introduzidas no procedimento DAGS, WinBUGS usa a verossimilhança para dados aumentados das variáveis observadas e as variáveis latentes auxiliares introduzidas (\mathbf{y}, \mathbf{Z}) (que são efeitos fixos e aleatórios na modelagem hierárquica)

para obter as distribuições posteriores para os parâmetros de interesse. Quando este é o caso, *DIC*s marginais tanto para as variáveis observadas (efeito fixo) quanto para as variáveis latentes auxiliares (efeitos aleatórios) são apresentadas na saída no pacote. Para uma comparação própria dos modelos propostos, nós consideramos o *DIC* marginal para as variáveis observadas dado que o foco da análise está em $p(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})$ e embora as variáveis aleatórias auxiliares sejam consideradas na estimação elas não são o objetivo da análise.

Tabela 2.3 *Resumo do coeficiente de variação (c.v.) das estimativas a posteriori dos a_j e b_j e Critério de Ajuste para o modelo probito-normal considerando diferentes prioris no exemplo do Teste de Matemática*

priori	c.v dos a_j		c.v dos b_j		Critério de Ajuste			
	média	dp	média	dp	Dbar	Dhat	ρ_D	DIC
A	0,590	0,344	-1,895	6,747	1445,00	1352,90	92,11	1537,11
B	0,698	0,549	-0,999	1,827	1453,16	1368,36	84,80	1537,96
C	0,890	1,033	-0,979	1,286	1461,09	1383,61	77,48	1538,57
D	0,901	1,090	-1,099	0,646	1453,22	1375,48	77,75	1530,97
E	0,532	0,114	-1,101	0,634	1450,14	1369,86	80,28	1530,43
F	0,481	0,111	-1,048	1,101	1446,60	1358,26	88,34	1534,94

A análise do *DIC* na Tabela 2.3 confirma que todas as prioris conduzem ao mesmo ajuste.

A Figura 2.2 apresenta comparações para prioris diferentes considerando o gráfico das médias versus desvios padrões dos coeficientes de variação a posteriori sobre as médias a posteriori do parâmetros de dificuldade e discriminação mostrados na tabela 3.2. Para os parâmetros de discriminação e dificuldade, as estimativas com diferentes prioris estão organizados no eixo horizontal da menor até a maior média do coeficiente da variação e o eixo vertical do menor até o maior desvio padrão do coeficiente de variação a posteriori. Claramente, prioris com grande precisão no parâmetro de discriminação, como é o caso das prioris truncadas *F* e *E*, obtém estimativas mais precisas. Este resultado sugere que prioris com hiperparâmetros que admitem valores negativos para os parâmetros de discriminação (média igual para zero nas prioris *C* e *D*) obtém estimativas mais imprecisas.

Com respeito ao parâmetro de dificuldade, as estimativas com diferentes prioris estão organizadas desde a priori menos precisa (em ordem, prioris *D* e *E*) até a mais precisa (menos vaga) (em ordem, prioris *A*, *B* e *C*). Uma grande valor de média e desvio padrão são obtidos para a priori *A* devido ao fato que a estimativa do coeficiente de variação para o parâmetro de dificuldade do Item 3 é grande. A imprecisão para o Item 3 segue do fato que as estimativas são próximas de zero e pequenos desvios padrões produzem um grande coeficiente de variação. Este

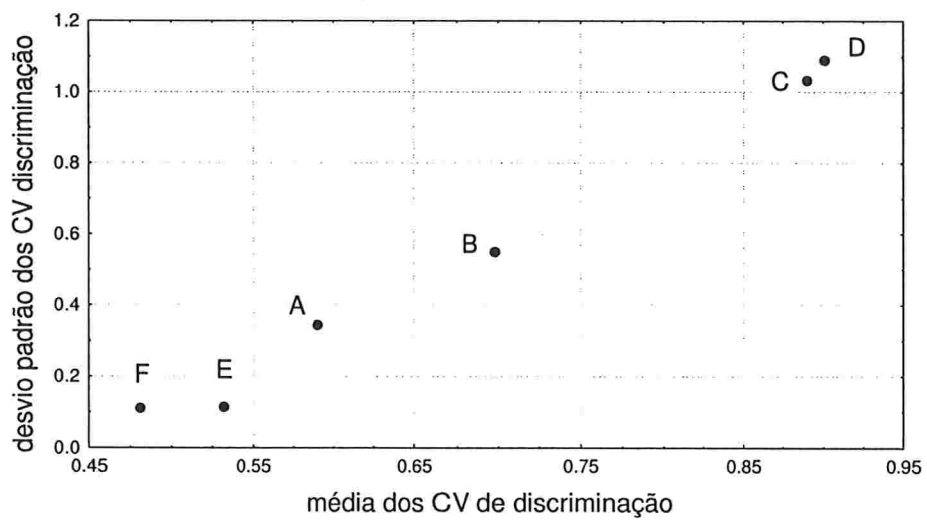
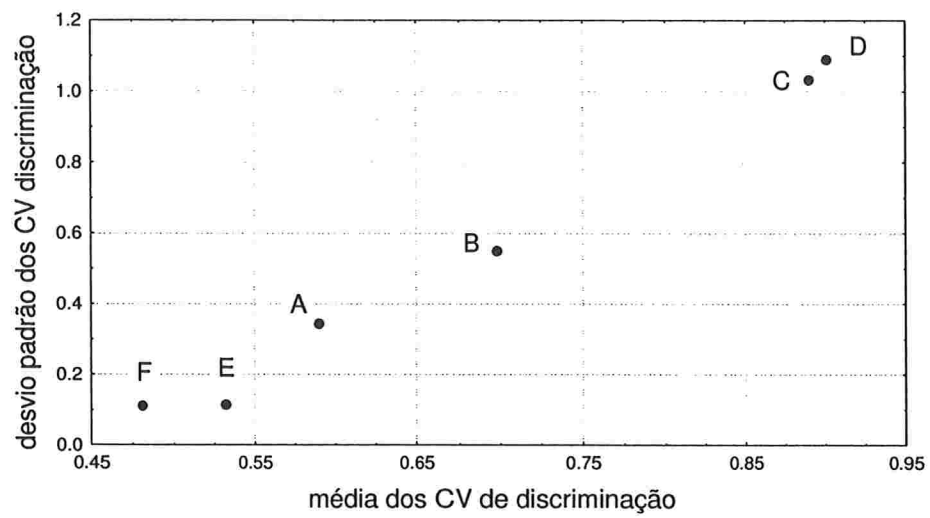


Figura 2.2 Comparação de prioris (A a F) do modelo probito-normal para o exemplo do Teste de Matemática usando resumos dos coeficiente de variação das estimativas a posteriori dos parâmetros a e b sob 14 itens

fato implica que para a priori A as estimativas para o parâmetro de dificuldade são próximas de zero.

Embora as prioris consideradas podem levar à uma análise a posteriori similar (ajuste), nós encontramos mais apropriada o uso das prioris E e F porque isto resulta em grande precisão para as estimativas dos parâmetros de discriminação e de dificuldade. Desde que prioris para o parâmetro de dificuldade com grande variância podem conduzir a distribuições a posteriori imprecisas nós consideramos a priori F como sendo a mais apropriada. Este fato é observado em Sahu (2002), em que outras escolhas para os hiperparâmetros foram também pesquisadas, e conclusões similares são observadas em nosso trabalho.

2.4.2 Inferência e interpretação do parâmetros

Estudos a seguir foram feitos considerando a seguintes distribuições a priori $a \sim N(1; 0.5)I(0,)$ e $b \sim N(0, 2)$ proposta em Sahu (2002). Nós consideramos um valor de descarte (burn-in) igual a 1000 e uma cadeia com 10000 iterações adicionais após do descarte com um valor de salto (thin) igual a 10 para reduzir a dependência das cadeias resultando num tamanho efetivo de amostra igual a 1000. Neste trabalho nós usamos cinco cadeias paralelas. A convergência foi diagnosticada usando a visualização do gráfico do traço da estatística de Gelman e Rubin (1992) apresentadas no pacote CODA (Best *et al.* 1997) dentro do programa R.

O tempo médio para o descarte tomou 90 segundos e para gerar 10000 iterações adicionais para cada cadeia foi de 957 segundos num processador AMD Duron(tm) com 112 MB RAM).

Figura 2.3 mostra o histórico da cadeia e a densidade empírica para os parâmetros a e b de um item do teste (item 12) e para a média e desvio padrão das variáveis latentes que descrevem a habilidades dos indivíduos. Note que as cadeias convergem a um valor estacionário e amostram densidades empíricas de uma moda so.

Estimativas dos parâmetros de discriminação e dificuldade do item são apresentadas na tabela 2.4. O item 11 é o mais discriminatório enquanto que o item 9 é o menos discriminatório. Também, o item 11 é o mais fácil e o item 12 é o mais difícil.

O item 11 diz: “Luiza, Dora e Maria compram certa quantidade de tela. Luiza comprou a metade de um metro, Dora comprou 75 centímetros e Maria comprou 50 centímetros. Qual delas comprou a mesma quantidade de tela?”. O item 12 diz: “Um recipiente recebe 4,5 litros de água por minuto. Quantos litros de água pode ter um recipiente depois de uma hora e meia?”. Por outro lado, o item 9 diz: “Resolva as seguintes operações em decimais: $0,75 - 0,2 + 1,2 - 0,30$ ”. Claramente, da leitura do enunciado dos itens, a habilidade (latente) requerida para resolver corretamente o item 11 é menor que a requerida para resolver o item 12. Para um valor dado de habilidade, um estudante têm maior probabilidade de sucesso com o item 11 do que com o item 12. Assim, de acordo ao modelo probito-normal, o item 11 é mais fácil do que o item 12. Também é importante notar a respeito do item 11 que um incremento da habilidade dos estudantes (conhecimento do significado do termo médio metro, por exemplo) se traduz em uma

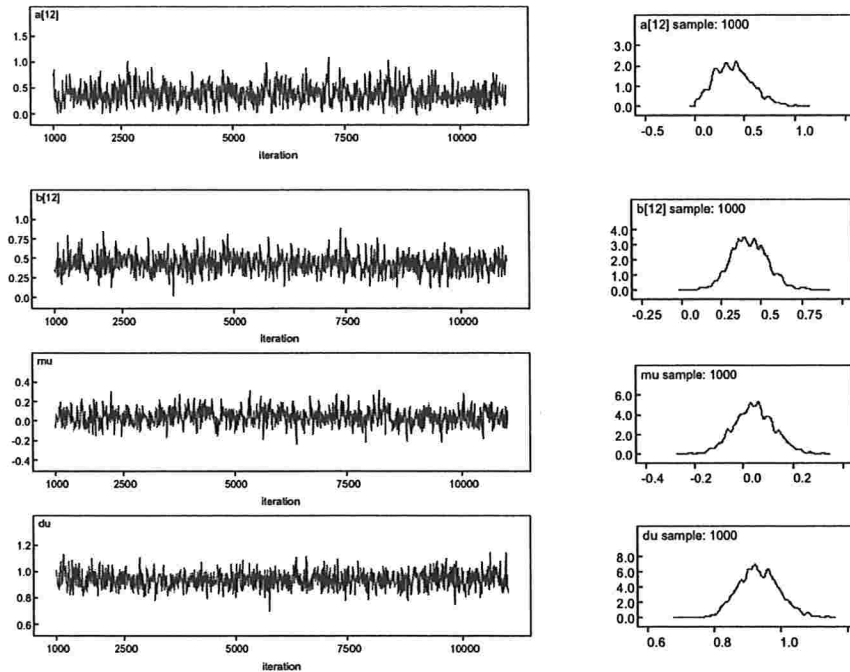


Figura 2.3 *Histórico e densidade da cadeia para os parâmetros a e b do item 12 ($a(12)$ e $b(12)$) e média (μ) e desvio padrão (du) das variáveis latentes u_j*

incremento da probabilidade de sucesso para este item. Por outro lado, com respeito ao item 9, pode ser observado que um incremento na habilidade do sujeito (conhecimento dos lugares decimais, por exemplo) não é diretamente traduzido em um incremento da probabilidade de responder corretamente o item. Assim, o item 11 parece ter um grande poder discriminatório porque permite distinguir melhor entre estudantes que tem ou não tem algum conhecimento específico a respeito dos ítems. Não em tanto, da tabela 2.4 se observa que os itens do teste não apresentam valores muito alto do parâmetro de discriminação.

Box-plots para os parâmetros dos ítems baseados na inferência posterior são apresentados na Figura 2.4 indicando a variação na distribuição a posterior para os parâmetros dos ítems. Este resultado obviamente é impossível de ser dado com a abordagem frequentista. Neste caso nós temos não so uma estimativa pontual dos parâmetros do modelo e sim a própria distribuição a posteriori o que claramente mostra as vantagens da estimação bayesiana para ter maior informação dos parâmetros. Para detalhes adicionais na interpretação dos parâmetros dos ítems veja Johnson e Albert (1999).

Tabela 2.4 *Resumo da inferência posterior para os parâmetros sob o modelo probito-normal no exemplo do Teste de Matemática*

		Estatísticas a posteriori				
		média	d.p.	$P_{2,5}$	mediana	$P_{97,5}$
Parâmetro de discriminação	a_1	0,54	0,24	0,13	0,52	1,09
	a_2	0,29	0,18	0,03	0,27	0,70
	a_3	0,55	0,24	0,16	0,53	1,09
	a_4	0,91	0,33	0,33	0,88	1,61
	a_5	0,49	0,24	0,09	0,48	1,02
	a_6	0,32	0,18	0,03	0,30	0,73
	a_7	0,88	0,34	0,29	0,84	1,60
	a_8	0,97	0,35	0,39	0,92	1,80
	a_9	0,20	0,14	0,01	0,18	0,51
	a_{10}	0,49	0,23	0,08	0,48	1,02
	a_{11}	1,35	0,41	0,64	1,32	2,18
	a_{12}	0,39	0,19	0,06	0,38	0,80
	a_{13}	0,45	0,23	0,05	0,43	0,93
	a_{14}	0,41	0,26	0,03	0,37	0,97
Parâmetro de dificuldade	b_1	-0,90	0,16	-1,25	-0,90	-0,60
	b_2	-1,09	0,15	-1,37	-1,08	-0,80
	b_3	-0,03	0,13	-0,27	-0,03	0,23
	b_4	-1,91	0,33	-2,62	-1,88	-1,36
	b_5	-1,23	0,18	-1,62	-1,22	-0,90
	b_6	0,37	0,11	0,16	0,37	0,61
	b_7	-1,82	0,33	-2,63	-1,78	-1,28
	b_8	-1,53	0,31	-2,30	-1,50	-1,03
	b_9	-0,80	0,13	-1,06	-0,80	-0,57
	b_{10}	-1,19	0,17	-1,55	-1,18	-0,87
	b_{11}	-2,31	0,46	-3,32	-2,25	-1,53
	b_{12}	0,43	0,12	0,21	0,42	0,67
	b_{13}	-1,01	0,16	-1,35	-1,00	-0,71
	b_{14}	-1,71	0,23	-2,20	-1,70	-1,32
Variável latente	média dos u_j	0,94	0,06	0,82	0,93	1,06
	d.p dos u_j	0,04	0,08	-0,12	0,04	0,21

Com respeito a habilidade latente, nos temos que a média a posteriori da média das habilidades dos estudantes é 0,94 a qual esta quase a um desvio padrão do média a priori fixada em 0. Isto pode indicar que a maioria dos estudantes apresentam alta habilidade o qual concorda

com escores observados que aparecem na figura 2.5. Assim habilidades estimadas sob o modelo probito normal estão diretamente associadas com os escores observados.

2.5 Discussão Final

Um dos objetivos deste capítulo foi contribuir na revisão da literatura associada ao modelo probito-normal na teoria da resposta ao item. Este modelo foi escolhido porque permite extensões para modelos de três parâmetros, modelos multidimensionais, modelos multinível, inclusão de variáveis preditivas e análise fatorial confirmatório. O capítulo considera dois procedimentos usando a metodologia MCMC para fazer a inferência para os parâmetros de itens e de habilidades no modelo. A metodologia fora usada num teste de matemática aplicado a 131 estudantes de sexta série de escolas peruanas. Foi feita uma análise de sensibilidade para checar a adequação do modelo que segue pelo uso de uma série de distribuições a priori também. Isto inclui a especificação de distribuições a priori vagas para os parâmetros de dificuldade e hiperparâmetros precisos para o parâmetros de discriminação. As prioris especificadas na Tabela 2.2 foram sugeridas em pesquisas anteriores por outros autores. As diferentes prioris consideradas levam a estimativas similares do DIC (descrito em Spiegelhalter *et al.*, 2002) levando a conclusão que o análise bayesiana para o conjunto de dados sob consideração não são sensíveis a escolha das prioris consideradas. Embora, a priori considerada por Sahu (2002), a qual especifica que $a_j \sim N(1; 0.5)I(0,)$ e $b_j \sim N(0, 2)$, $j = 1 \dots$, parece ser, para nós, mais adequada. No Capítulo 4 (veja também Bazán *et al.*, 2004) nós propomos o uso de uma família mais geral de funções de ligação $F_{ij}(\cdot)$ a qual também pode acomodar distribuições assimétricas para as habilidades na população que podem ser aplicados neste conjunto de dados, já que a distribuição dos escores assimétrica poderia ser melhor ajustada com um modelo que tome em conta a assimetria.

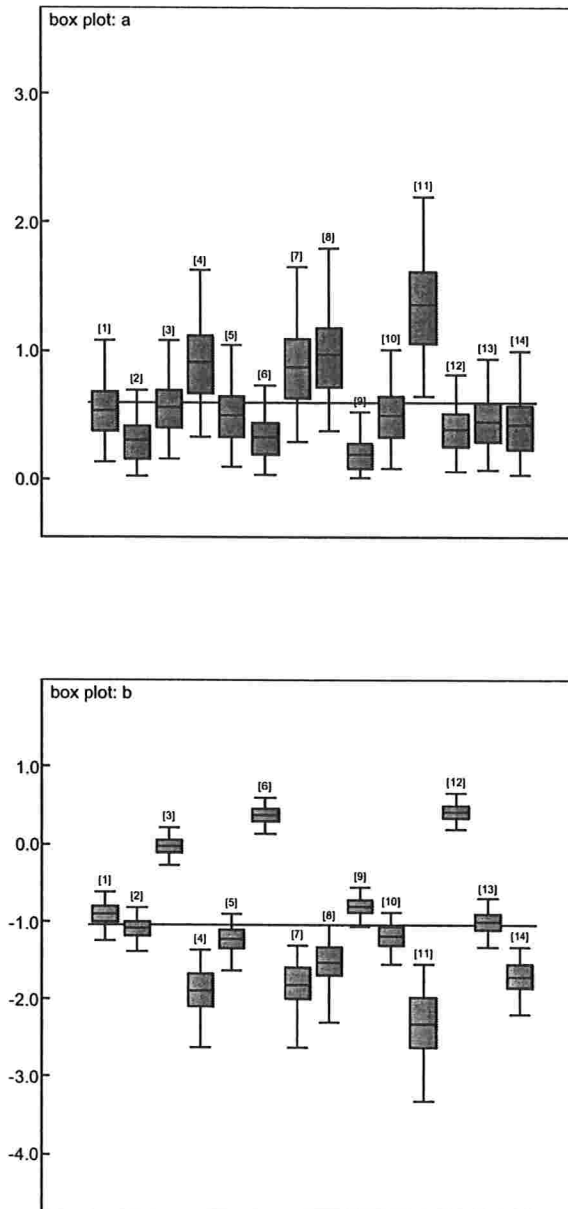


Figura 2.4 Diagrama de caixas (“box-plot”) para parâmetros de discriminação (a) e dificuldade (b) dos itens no modelo probito-normal

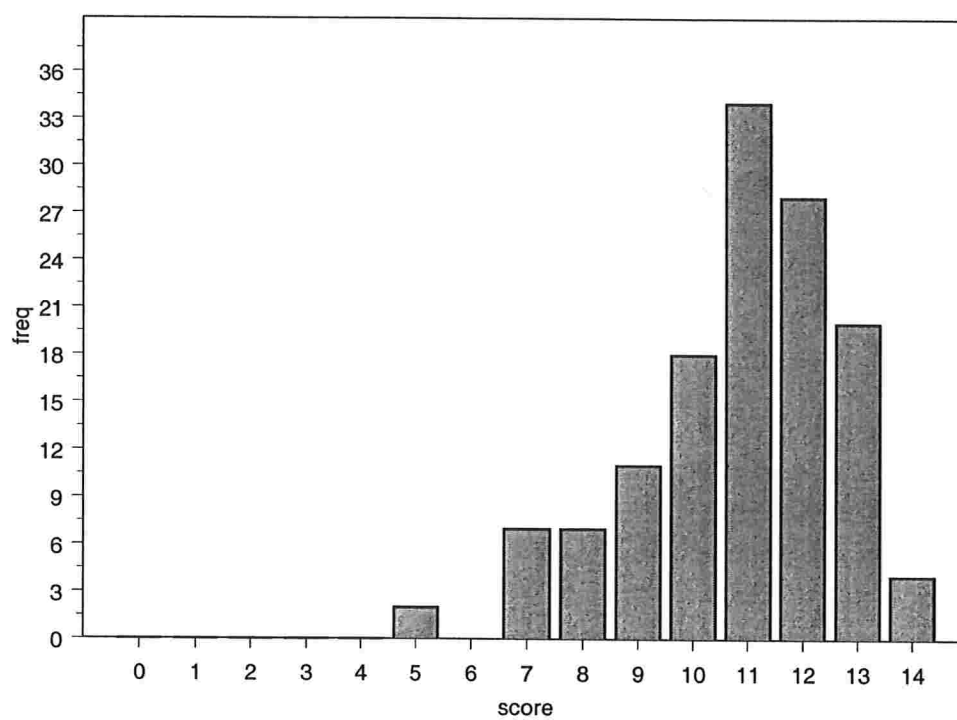


Figura 2.5 *Histograma dos escores do teste de Matemática de 14 itens para 131 estudantes*

Capítulo 3

Uma função de ligação probito assimétrica generalizada para resposta binária

3.1 Introdução

Tradicionalmente, os modelos utilizados para dados binários são os modelos probito e logito. Em ambos os casos, a função de ligação é simétrica. Entretanto, quando a probabilidade de sucesso se aproxima de 0 a uma taxa diferente do que se aproxima de 1, ligações simétricas podem ser inadequadas como é discutido, por exemplo, em Chen *et al.* (1999). Neste caso, ligações assimétricas podem ser convenientes. Varias ligações assimétricas paramétricas tem sido desenvolvidas na literatura e vários exemplos são listados em diferentes livros texto.

Neste capítulo propomos uma família paramétrica de ligações assimétricas bastante flexível. Para a construção da nova ligação, a fda da distribuição normal assimétrica (Dalla Valle, 2004) é considerada. Antes é recomendável revisar o apêndice A para uma apresentação das propriedades conhecidas desta distribuição e as novas propriedades derivadas nesta teses assim como casos particulares da normal assimétrica que serão usadas a seguir.

Neste capítulo introduzimos a função de ligação probito assimétrica generalizada para o modelo de resposta binária a qual tem como casos particulares a ligação probito assimétrica dada por Chen *et al.* (1999) baseada na normal assimétrica *CDS*, a função de ligação probito dada por Bazán *et al.* (2004a) a qual se baseia na normal assimétrica *BBB* e a usual ligação probito. Apresentamos algumas propriedades desta ligação assimétrica e também condições para a existência das estimativas de máxima verossimilhança e da distribuição a posteriori para o parâmetro de regressão. Além disso, é desenvolvido o algoritmo MCMC e, finalmente, para um conjunto de dados bastante conhecido na literatura, é desenvolvida uma comparação de modelos considerando diferentes funções de ligação.

O capítulo está organizado da seguinte maneira. A Seção 3.2 revisa as ligações para resposta binária na literatura. A Seção 3.3 apresenta o modelo probito assimétrico generalizado. A Seção 3.4 são discutidos resultados principais a respeito da existência de estimativas de máxima verossimilhança e propriedades da distribuição posterior dos parâmetros do modelo quando a priori uniforme imprópria é escolhida são discutidas. Aumento de dados é considerada na Seção 3.5, e uma metodologia da estimação Bayesiana é desenvolvida na Seção 3.6. Na Seção 3.7 um exemplo de aplicação é considerado.

3.2 Ligações para resposta binária

Considere $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ um vetor $n \times 1$ de n variáveis aleatórias independentes, $\mathbf{x}_i = (x_{i1}, \dots, x_{in})'$ um vetor $k \times 1$ de covariáveis. Seja \mathbf{X} a matriz de planejamento $n \times k$ com linhas \mathbf{x}_i' , e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ um vetor $k \times 1$ de coeficientes de regressão. Suponha que $y_i = 1$ com probabilidade p_i e $y_i = 0$ com probabilidade $1 - p_i$. Numa modelagem para dados binários, é usual assumir que

$$p_i = F(\eta_i) = F(\mathbf{x}_i' \boldsymbol{\beta}), \quad i = 1, \dots, n, \quad (3.1)$$

na qual $F(\cdot)$ denota uma função de distribuição acumulada (fda). A inversa F^{-1} é tipicamente denominada função de ligação e $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$, i -ésimo predictor linear. Quando F é a fda de uma distribuição simétrica a ligação resultante é simétrica e p_i tem uma forma simétrica em torno de $p_i = 0.5$. Exemplos são obtidos quando F é a fda na família das distribuições elípticas (Arellano-Valle, 1994) como a Normal Padrão, Logística, T-student, Exponencial Dupla ou distribuição de Cauchy (veja-se Albert e Chib, 1993).

3.2.1 Ligações assimétricas

Existem diversas propostas para se obter ligações assimétricas, a saber:

- (i) Considerando, em (3.1), que F é uma fda de uma distribuição assimétrica;
- (ii) Considerando, em (3.1), uma modificação do predictor linear η_i dada por $\eta_i = m(\eta_i, \lambda)$, onde $m(\cdot)$ é uma função não linear e contínua e λ está associado ao grau de assimetria;
- (iii) Considerando, em (3.1), que $F(\cdot)$ é uma fda na classe de misturas de escala de distribuições elípticas.

Na Figura 3.2.1 temos um exemplo de função de ligação não simétrica. Ligações assimétricas do tipo (i) são obtidas, por exemplo, quando distribuições assimétricas tais como a Gumbel (que define a ligação loglog) e a Weibull são consideradas. Entretanto, nestas distribuições não existe um parâmetro que modele a assimetria, e assim nenhuma relação é estabelecida entre as ligações simétricas e assimétricas. Ou seja, não se pode, por exemplo, obter as simétricas a partir das assimétricas, portanto, as simétricas não podem ser consideradas como um caso particular das assimétricas.

Exemplos do caso (ii) são dados em Prentice (1976) e Czado e Santner (1992a,b) que consideram a distribuição padrão de Burr; Guerrero e Johnson (1982) e Stukel (1988) usam a distribuição logística padrão e Czado (1994) que usa a distribuição normal padrão mas com $m(\cdot)$ modificado. Assim p_i é obtida considerando $p_i = F(m(\eta_i, \lambda))$ para algum limite de variação para η_i (por exemplo para $\eta_i \geq 0$) e $p_i = F(\eta_i)$ para outro intervalo de variação de η_i (por exemplo para

$\eta_i < 0$). Somente no caso da distribuição de Burr, $p_i = F(m(\eta_i, \lambda))$ não é dada por uma função particionada e não é necessário partições de η_i em intervalos na definição de $m(\cdot)$. Outros casos são obtidos quando o preditor linear é substituído por uma expressão polinomial, geralmente quadrática ou cúbica (veja, por exemplo, Collet, 2003).

Outras famílias mais de ligações são obtidas quando F é escolhida na classe de funções de distribuição acumulativas de mistura de escalas elípticas $\mathcal{F} = \{F(\cdot) = \int_{[0, \infty]} H(\cdot|\lambda)dG(\lambda)\}$, onde G é uma fda em $[0, \infty]$ e H é uma fda elíptica (veja Basu e Mukhopadhyay, 2000, Haro-López *et al.*, 2000 e Arellano-Valle, 1994).

Nós revisamos na seção seguinte as formas de obter ligações probito assimétricas consideradas na literatura que aparentemente levam a diferentes modelos.

3.2.2 Ligações probito assimétricas

A primeira formulação de uma ligação probito assimétrica foi dada por Czado (1994). Esta ligação é obtida usando a forma (ii) de obter ligações assimétricas, com F sendo a fda da normal padrão. Czado (1994) estuda o caso de uma particular $m(\cdot)$. Ainda que Czado (1994) tenha usado uma particular função $m(\cdot)$, em Czado (1997) é apresentada uma lista de várias funções $m(\cdot)$ que podem ser usadas para definir ligações probito assimétricas. Uma crítica a esta construção é o fato de estas ligações assimétricas serem limitadas desde que $m(\cdot)$ é particionada para diferentes intervalos de variação de η_i e todas as possibilidades precisam ser checadas para inferir o sinal do parâmetro de assimetria (veja por exemplo Czado, 1994, Czado, 2000).

Uma formulação mais recente de ligação probito assimétrica foi proposta em Chen *et al.* (1999), Chen *et al.* (2001) e Chen (2004). Eles propuseram uma classe de ligações assimétricas quando se considerada:

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

onde

$$z_i = \eta_i + \varepsilon_i, \quad (3.3)$$

e

$$\varepsilon_i = \lambda v_i + u_i, \quad v_i \sim G, \quad u_i \sim H, \quad (3.4)$$

Assim, a ligação proposta por Chen *et al.* (1999) é dada por

$$p_i = \int_{-\infty}^{\infty} H(\eta_i + \lambda v_i)g(v_i)dv_i. \quad (3.5)$$

onde v_i e u_i são independentes, G é a fda de uma distribuição assimétrica, H é a fda de uma distribuição simétrica e g a função de densidade associada a G .

A expressão (3.4) corresponde a representação estocástica dada por Sahu *et al.* (2003) para a construção de distribuições assimétricas, em que $\lambda \in \mathbb{R}$ é um parâmetro de assimetria. Na formulação de Chen *et al.* (1999), p_i não tem uma expressão analítica. Assim, para estimação dos

parâmetros do modelo, Chen *et al.* (1999) apresentam uma versão de verossimilhança aumentada considerando (3.5).

Quando H é a fda da Normal e G é a fda da normal positiva, a fda de z_i é a fda da distribuição normal assimétrica dada por Sahu *et al.* (2003) e Chen *et al.* (1999), isto é, usando a notação dada no Apêndice A e em Dalla Valle (2004), $z_i \sim NA(\theta)$ com $\theta = (\eta_i, 1 + \lambda^2, \lambda)$ com locação η_i , escala $1 + \lambda^2$ e parâmetro de assimetria λ . Neste caso, como é visto no Apêndice A, para qualquer normal assimétrica existe uma expressão analítica para sua fda e conseqüentemente para p_i , fato que não foi notado por Chen *et al.* (1999). Esta ligação probito assimétrica nós denotaremos por *CDS probito assimétrico* ou simplesmente *CDS pa*.

Outra formulação para uma ligação probito assimétrica é considerar diretamente F como fda da distribuição normal assimétrica padrão (Azzalinni, 1985) a qual será usada no desenvolvimento dos capítulos 4 e 5. Neste caso, como é visto no Apêndice A, temos que sua fda é dada por

$$p_i = F_\lambda(\eta_i) = 2\Phi_2\left(\left(\begin{array}{c} \eta_i \\ 0 \end{array}\right); \left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & -d \\ -d & 1 \end{array}\right)\right). \quad (3.6)$$

onde $F_\lambda(\cdot)$ denota a fda da distribuição normal assimétrica com parâmetro de assimetria $\lambda \in \mathbb{R}$, e $\Phi_2(\cdot)$ denota a função de distribuição da distribuição normal bivariada com coeficiente de correlação $-d$, onde $d = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$ (veja notação no Apêndice A). Esta ligação probito assimétrica será denotada por *BBB probito assimétrica* ou simplesmente *BBB pa*. Para estimação, a verossimilhança segue a verossimilhança usual no modelo Bernoulli. No entanto, quando uma estrutura linear latente é considerada como a dada em (3.2)-(3.4) em que (3.4) é substituída pela representação estocástica de Henze (1986), como visto no Apêndice A obtemos que z_i é a fda de uma distribuição normal assimétrica diferente da dada por Chen *et al.* (1999). Isto é, $z_i \sim NA(\theta)$ onde $\theta = (\eta_i, 1, -\lambda)$, com locação η_i , escala 1 e parâmetro de assimetria $-\lambda$.

Um aspecto comum a *CDS pa* e *BBB pa* é que o parâmetro de assimetria está associado com a distribuição de $F(\cdot)$ e é independente do predictor linear. No entanto, a formulação de ambos modelos é diferente. Na ligação *CDS pa* sua formulação é obtida considerando uma estrutura linear latente e nenhuma forma fechada como em (3.1) é dada para p_i em (3.5). Na ligação *BBB pa* uma estrutura linear latente não é necessária para a definição do modelo pois é suficiente para sua formulação a escolha de uma nova distribuição para F . Embora as formulações de ligações probito assimétricas *CDS pa* e *BBB pa* sejam diferentes, veremos que se tratam de casos particulares de um modelo mais geral.

A ligação probito assimétrica generalizada para o modelagem de resposta binária considerada neste capítulo tem como casos particulares o *CDS pa*, *BBB pa* e a usual ligação probito. Também, outras ligações probito assimétricas podem ser derivadas. Esta abordagem considera o seguinte: a) p_i é obtida quando considerada a fda de uma função de distribuição avaliada no predictor linear, b) o parâmetro de assimetria está associado com a distribuição de $F(\cdot)$ e é independente do predictor linear, e c) a estrutura linear latente não é necessária para a formulação do modelo.

A ligação assimétrica proposta introduz um parâmetro que controla a taxa de crescimento (ou decréscimo) da probabilidade de sucesso (falha) da resposta binária e outros parâmetros de locação e escala desta probabilidade.

3.3 Modelo probito assimétrico generalizado

3.3.1 Modelo

Um modelo probito assimétrico generalizado é obtido considerando F em (3.1) como a fda da distribuição normal assimétrica com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$, em que μ é um parâmetro de locação, σ^2 é um parâmetro de escala e λ é um parâmetro de assimetria. Assim, usando a notação introduzida no Apêndice A, nós temos que

$$p_i = F_{\theta}(\eta_i), \quad i = 1, \dots, n \quad (3.7)$$

A formulação considerando uma distribuição normal assimétrica com parâmetros de locação, escala e assimetria constitui uma família de modelos que leva aos casos particulares vistos na Seção 3.2.2.

Proposição 3 Considerando a ligação probito assimétrica generalizado temos que

- Se $\theta = (0, 1, 0)$ segue a ligação probito.
- Se $\theta = (0, 1 + \lambda^2, -\lambda)$ segue a ligação *CDS pa*.
- Se $\theta = (0, 1, \lambda)$ segue a ligação *BBB pa*.

Prova. Veja o Apêndice A e considere as proposições A.1 e A.2 e a definição da função de ligação dos modelos *CDS pa* e *BBB pa* como explicada na seção anterior. \square

A formulação do modelo probito assimétrico generalizado essencialmente implica que o modelo probito é caso particular do modelo probito assimétrico. Para ver isto no caso particular de *BBB pa*, nós apresentamos várias relações no Apêndice C.

A Figura 3.1 mostra diferentes curvas de probabilidade dos modelos *CDS* e *BBS pa* usando escalas diferentes para η_i . Note que o *CDS pa* apresenta probabilidade de sucesso que cresce lentamente em relação a ligação *BBB pa*. Esta primeira, pode ser mais adequada se o intervalo de variação de η_i for mais amplo.

Claramente, outras ligações probito assimétricas podem ser derivadas a partir de (3.7) quando consideramos outras escolhas para μ e σ^2 . Em geral é mais conveniente considerar $\mu = 0$ e usar σ^2 para o controle do crescimento da probabilidade de sucesso para um mesmo valor de λ e para maior flexibilidade.

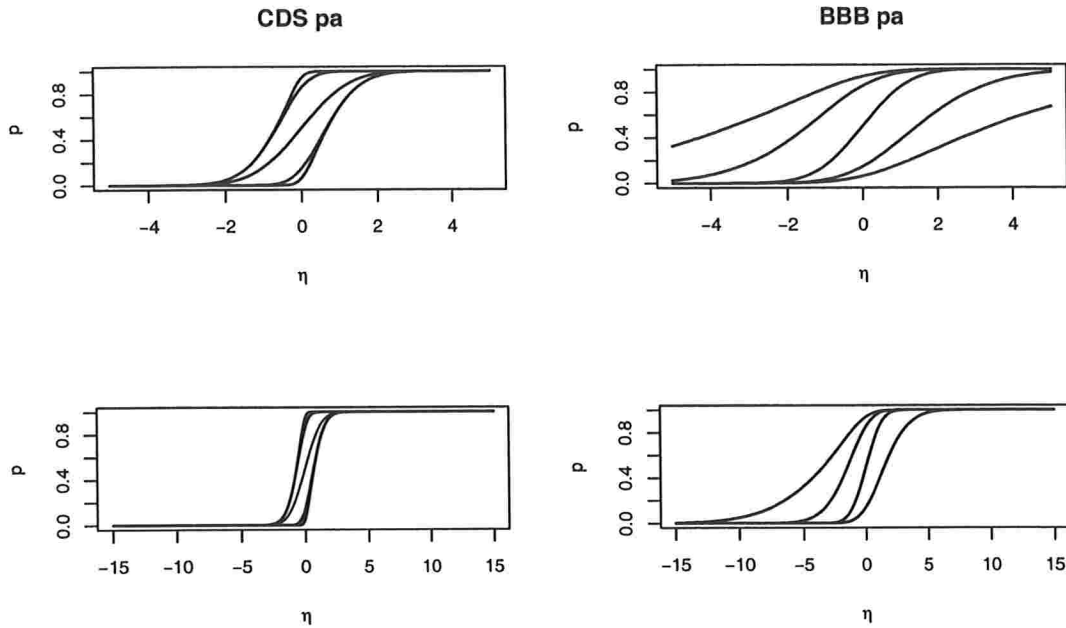


Figura 3.1 Curvas de probabilidade dos modelos CDS pa e BBB pa considerando diferentes valores para $\lambda = -5, -2, 0, 2, 5$ (nesse ordem) para escalas diferentes para $\eta = \mathbf{x}'\boldsymbol{\beta}$

3.3.2 Verossimilhança

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

$$L(\boldsymbol{\beta}, \theta | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n [F_{\theta}(\eta_i)]^{y_i} [1 - F_{\theta}(\eta_i)]^{1-y_i}. \quad (3.8)$$

Considerando a proposição A.1 no Apêndice A, nós temos que para $\theta = (\mu, \sigma^2, \lambda)$

$$F_{\theta}(\eta_i) = 2\Phi_2\left(\left(\begin{array}{c} \eta_i \\ 0 \end{array}\right); \left(\begin{array}{c} \mu \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma & -d \\ -d & 1 \end{array}\right)\right) \quad (3.9)$$

e (3.8) pode ter outra versão baseada na fda da distribuição normal bivariada. O uso do algoritmo para a fda da normal assimétrica em (3.7) esta baseado na função de Owen (Azzalini, 1985 e Dalla Valle, 2004), e são disponíveis em programas R e Matlab. Entretanto, vários algoritmos computacionais eficientes são disponíveis para calcular integrais relacionadas com a distribuição normal bivariada, veja, por exemplo, Genz (1992, 1993), e a versão usando (3.9), que algumas vezes podem ser mais convenientes.

Além disso, considerando a proposição 2 no Apêndice A, nós temos que para $\theta = (\mu, \sigma^2, \lambda)$

$$F_{\theta}(\eta_i) = \int_0^{\infty} \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v_i\right) g(v_i) dv_i \quad (3.10)$$

onde $\Phi(\cdot)$ denota a fda da distribuição normal padrão e $g(v) = \frac{2}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} I(0, \infty)$ é a função de densidade de probabilidade (fdp) da distribuição Normal positiva ou Half Normal (veja Apêndice A).

Como conseqüência, a função de verossimilhança para os modelos probito assimétricos generalizados pode ser escrita (Chen *et al*, 1999) como

$$L(\beta, \theta | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \int_0^{\infty} \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v_i\right)^{y_i} \left[1 - \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v_i\right)\right]^{1-y_i} g(v_i) dv_i. \quad (3.11)$$

Uma versão aumentada da função de verossimilhança é obtida considerando-se as variáveis auxiliares v_i , e é dada por

$$L(\beta, \theta | \mathbf{y}, \mathbf{X}, \mathbf{v}) = \prod_{i=1}^n \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v_i\right)^{y_i} \left[1 - \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v_i\right)\right]^{1-y_i} g(v_i). \quad (3.12)$$

Considerando a Proposição 3, é fácil obter as versões da função de verossimilhança para os modelos *BBB* e *CDS pa* como casos particulares de (3.8), (3.9) e (3.12).

3.4 Resultados principais para prioris impróprias

Note que no modelo probito assimétrico generalizado o parâmetro θ está associado a escolha da função de ligação mas β representa um parâmetro com significado específico na estrutura escolhida e eles podem ser considerados como inerente aos dados observados e não dependentes da definição de um modelo específico (para uma maior discussão a respeito desta distinção de parâmetros, ver Taylor e Siqueira (1996) e Drapper (1995)). Portanto, dois cenários podem ser considerados. O primeiro cenário é que ambos parâmetros β e θ são estimados; no segundo cenário só β varia e θ é fixado com valor θ_0 “verdadeiro”. Semelhante a Taylor e Siqueira (1996), nós podemos referir a esses dois cenários como abordagem não condicional e abordagem condicional, respectivamente.

Na abordagem condicional, claramente, como no caso do modelo probito, calcular as estimativas de máxima verossimilhança (MLE) para β usando as versões da função de verossimilhança consideradas na Seção 3.3 é simples. Também, a inferência bayesiana pode ser facilmente formulada após a especificação de uma distribuição a priori para β . Mas condições devem ser impostas sob a matriz de planejamento para a existência das estimativas de MV e para a existência da distribuição a posteriori de β sob uma priori uniforme imprópria. Alguns resultados sobre a

existência da estimativa de máxima verossimilhança e da posteriori quando prioris impróprias são consideradas, são apresentados a seguir. As provas são baseadas em resultados apresentados em Chen and Shao (2000).

3.4.1 Existência da posteriori e da estimativa de máxima verossimilhança na abordagem condicional

Proposição 4. Considere a abordagem condicional para a ligação assimétrica generalizada, isto é, em que θ é conhecido ou fixado e só β é o parâmetro de interesse, e seja $t_i = 1$ se $y_i = 0$ e $t_i = -1$ se $y_i = 1$, \mathbf{X} é a matriz de planejamento $n \times k$ conhecida com linhas \mathbf{x}'_i e defina \mathbf{X}^* como a matriz com linhas $t_i \mathbf{x}'_i$. Sob, as seguintes condições:

(C1) \mathbf{X} é de posto completo;

(C2) Existe um vetor positivo $\mathbf{a} = (a_1, \dots, a_n)'$ tal que $\mathbf{X}^* \mathbf{a} = \mathbf{0}$,

então,

- Para uma priori uniforme imprópria para β , isto é, $\pi(\beta) \propto 1$ a distribuição posterior de β é própria, isto é, $\int_{R^k} L(\beta|\mathbf{y}) d\beta < \infty$;
- A estimativa de máxima verossimilhança (MV) de β existe.

Prova. Quando θ é conhecido, a verossimilhança dada em (3.8) se reduz a $L(\beta|\mathbf{y}, \mathbf{X}, \theta)$ que por simplicidade escrevemos $L(\beta|\mathbf{y})$, e considerando a priori $\pi(\beta) \propto 1$ então, para que a posteriori seja própria requeremos que $\int_{R^k} L(\beta|\mathbf{y}) d\beta < \infty$. Se (C1) e (C2) se satisfazem, pelo Teorema 2.1 de Chen e Shao (2000) tais condições são suficientes desde que a normal assimétrica é uma função de distribuição contínua que satisfaz $\int_{-\infty}^{\infty} e^{t\tau} dF_{\theta}(\tau) = E(e^{tR}) = M_R(t) < \infty$, e $E(|u^k|) = \frac{d^k M_R(t)}{dt^k} |_{t=0}$. Isto é tem momentos finitos, como pode ser verificado considerando a expressão dada no Apêndice A para sua função geradora de momentos.

A existência da estimativa de MV esta garantida desde que a fda da normal assimétrica é contínua e pelo Teorema 3.1 dados em Chen e Shao (2000) (C1) e (C2) são suficientes. \square

3.4.2 Existência da posteriori e da estimativa de máxima verossimilhança na abordagem não condicional

Na abordagem não condicional, calcular as estimativas de MV usando as versões da função de verossimilhança dada na Seção 3.3 não é simples e novas condições para a existência das estimativas de MV são necessárias. Além disso, para a especificação de uma distribuição a priori uniforme imprópria para β e θ é necessário ter outras condições para a existência da distribuição

3.5 Verossimilhanças aumentadas para o modelo probito assimétrico generalizado 84

a posterior de β como, por exemplo, as dadas em Chen *et al.*, (1999). Em geral μ e σ^2 devem ser fixados como acontece nas ligações *CDS* e *BBB probito assimétricas*, sendo λ o único parâmetro de interesse associado com a escolha da ligação probito assimétrica generalizada. Considerando este fato, nós temos a seguinte proposição.

Proposição 5. *Considere a abordagem não condicional para a ligação probito assimétrica generalizada, isto é, em que θ e β são parâmetros de interesse, e seja $S_0 = \{i : y_i = 0\}$, $S_1 = \{j : y_j = 1\}$, $m_0 = \text{card}(S_0)$, $m_1 = \text{card}(S_1)$ e X_0 e X_1 são matrizes com linhas x'_i , $i \in S_0$ e x'_j , $j \in S_1$. Sob as seguintes condições:*

(C1') X_0 e X_1 são de posto completo;

(C2') Existem vetores positivos $\mathbf{a}_0 = (a_{01}, \dots, a_{0n})' \in R^{m_0}$ e $\mathbf{a}_1 = (a_{11}, \dots, a_{1n})' \in R^{m_1}$ tal que $X'_0 \mathbf{a}_0 = \mathbf{0}$ e $X'_1 \mathbf{a}_1 = \mathbf{0}$

então,

- Para uma priori uniforme imprópria conjunta para β e λ , isto é, $\pi(\beta, \lambda) \propto 1$ a distribuição posterior conjunta de β e λ é própria, isto é, $\int_{-\infty}^{\infty} \int_{R^k} L(\beta, \lambda | y) d\beta d\lambda < \infty$;
- A estimativa de máxima verossimilhança (MV) de β existe.

Prova. Desde que para $\theta = (\mu, \sigma^2, \lambda)$ tem-se especificada uma priori uniforme imprópria apenas para λ pois μ e σ^2 se assumem conhecidos, o problema de garantir a posteriori própria quando especificada também uma priori uniforme imprópria para β já fora abordado por Chen *et al.* (1999) para a normal assimétrica *CDS*. Para a prova no caso da normal assimétrica geral é conveniente usar a versão da função de verossimilhança dada em (3.11) onde $g(\cdot)$ é a fdp da normal positiva que tem suporte nos reais positivos. Com isto a proposição é análoga aos Teoremas 2 e 5 dado por eles. As condições indicadas no Teorema 5 de Chen *et al.* (1999) são satisfeitas desde que a normal assimétrica es uma função de distribuição contínua e assim a existência da estimativa de máxima verossimilhança está garantida. Além disso como fora indicado na prova da proposição 4 a fda de uma distribuição normal assimétrica tem momentos finitos e então, se satisfazé também a outra condição indicada no Teorema 2 e pelo tanto a posteriori com priori uniforme imprópria em λ e β é própria. \square

3.5 Verossimilhanças aumentadas para o modelo probito assimétrico generalizado

Semelhante a versão de verossimilhança aumentada para o modelo probito, introduzida por Albert e Chib (1993) e vista no capítulo 2, e diferente da proposta por Chen *et al.* (1999)

3.5 Verossimilhanças aumentadas para o modelo probito assimétrico generalizado 85

apresentada em 3.2.2, nós propomos nesta Seção uma verossimilhança aumentada para o modelo probito assimétrico generalizado. Considerando-se variáveis latentes auxiliares com distribuição normal assimétrica padrão. É fácil mostrar que a ligação probito assimétrico é equivalente a considerar que

$$y_i = I(Z_i > 0) = \begin{cases} 1, & Z_i > 0; \\ 0, & Z_i \leq 0, \end{cases} \quad i = 1, \dots, n \quad (3.13)$$

onde $Z_i \sim NA(\theta)$, com $\theta = (\mu + \eta_i, \sigma^2, -\lambda)$, e $I(\cdot)$ como já fora definida antes, é a função indicadora. Claramente, no caso especial $\theta = (0, 1, 0)$ temos o resultado de Albert e Chib (1993). A variável latente Z_i é introduzida para evitar trabalhar com verossimilhanças de tipo de Bernoulli e esta representação mostra uma estrutura linear latente, a qual produz um modelo equivalente ao modelo probito assimétrico generalizado. Note que o sinal do parâmetro de assimetria para a variável latente auxiliar é oposto ao do parâmetro de assimetria especificado em (3.7).

Então, a *função de verossimilhança de dados aumentados* para o modelo probito assimétrico generalizado é dada por

$$L(\beta, \theta | Z, y) = \prod_{i=1}^n f_{\theta}(z_i) I(Z_i, y_i), \quad (3.14)$$

em que $I(Z_i, y_i) = I(Z_i > 0)I(y_i = 1) + I(Z_i \leq 0)I(y_i = 0)$, $i = 1, \dots, n$.

No Apêndice C são apresentadas algumas relações entre esta verossimilhança aumentada para a ligação probito assimétrica e a verossimilhança aumentada para a ligação probito.

A verossimilhança dada em (3.14) depende da fdp da normal assimétrica o que pode ser restritivo do ponto de vista inferencial em alguns casos. No entanto, é possível obter uma verossimilhança de dados aumentados alternativa para o modelo probito assimétrico generalizado. Primeiro considere a seguinte representação linear latente para z_i :

$$Z_i = \eta_i + e_i, \quad e_i \sim NA(\theta^*), \quad i = 1, \dots, n \quad (3.15)$$

em que $\theta^* = (\mu, \sigma^2, -\lambda)$. Além disso, para e_i , considere a representação estocástica para a distribuição normal assimétrica padrão (Henze, 1986) como dado no Apêndice A, isto é

$$e_i = \mu - \sigma(dV_i + (1 - d^2)W_i), \quad i = 1, \dots, n \quad (3.16)$$

em que $V_i \sim HN(0, 1)$, é a distribuição normal positiva, e $W_i \sim N(0, 1)$, a distribuição normal padrão. Segue que a distribuição condicional $e_i | V_i = v_i$ é uma distribuição normal com média $\mu - \sigma dv_i$ e variância $(1 - d^2)\sigma^2$. De acordo com este resultado, simulação de Z_i na estrutura linear (3.15) pode ser feita em duas etapas. Primeiro, simule $V_i \sim HN(0, 1)$ e então simule $Z_i^* = Z_i | V_i \sim N(\eta_i + \mu - \sigma dv_i, (1 - d^2)\sigma^2)$.

Portanto, a *função de verossimilhança de dados aumentados alternativa* para o modelo probito

assimétrico é dado por

$$L(\beta, \theta | Z, y) = \prod_{i=1}^n \phi(z_i; \eta_i + \mu - \sigma d v_i, (1 - d^2)\sigma^2) \phi(v_i; 0, 1) I(Z_i, y_i) I(v_i > 0), \quad (3.17)$$

em que $I(Z_i, y_i) = I(Z_i > 0)I(y_i = 1) + I(Z_i \leq 0)I(y_i = 0)$, $i = 1, \dots, n$.

A verossimilhança aumentada alternativa é acima é muito conveniente para estimação e é fácil obter as versões da função de verossimilhança aumentada e aumentada alternativa para os modelos *BBB* e *CDS* pa substituindo os valores correspondentes de θ especificados na proposição 3. Também note que o erro e_i introduzido em (3.16) são residuais latentes i.i.d e podem ser usados para a verificação do modelo (veja Albert e Chib, 1995). A distribuição a priori dos resíduos e_1, \dots, e_n , fornece uma base conveniente com a qual comparar a distribuição a posteriori. Neste caso, usando (3.16), alguma álgebra leva a

$$\pi(e_i | \beta, \theta, y) = \begin{cases} \frac{f_{\theta^*}(e_i) I(e_i > -\mu - \eta_i)}{F_{\theta^*}(\eta_i)}, & \text{if } y_i = 1, \\ \frac{f_{\theta^*}(e_i) I(e_i < -\mu - \eta_i)}{F_{\theta^*}(-\eta_i)}, & \text{if } y_i = 0, i = 1, \dots, n. \end{cases} \quad (3.18)$$

em que $\theta^* = (\mu, \sigma^2, -\lambda)$, $f_{\theta^*}(\cdot)$ e $F_{\theta^*}(\cdot)$ são a função de densidade e a fda da normal assimétrica correspondente.

3.6 Análise Bayesiana

3.6.1 Especificação de prioris

No contexto bayesiano é requerido especificar uma distribuição a priori para β digamos, $\pi(\beta)$ e para θ , digamos, $\pi(\theta)$. Por facilidade, é comum considerar

$$\pi(\beta, \theta) = \pi_1(\beta) \pi_2(\theta), \quad (3.19)$$

isto é, β e θ independentes a priori. Procedimento Bayesianos podem então ser implementados considerando as funções de verossimilhança dadas em (3.11), (3.12), (3.14) ou (3.17).

Considerando uma estrutura de independência a priori como em (3.19) podemos escolher as prioris para $\pi(\beta)$ usualmente consideradas no modelo probito incluindo prioris impróprias. Assim, podemos assumir como priori para β , uma distribuição normal, prioris impróprias do tipo uniforme para β ($\pi(\beta) = 1$) ou ainda prioris de Jeffreys. Veja, por exemplo, Zellner e Rosi (1984), Johnson e Albert (1999) e Ibrahim and Laud (1991).

De um modo geral nós consideramos para θ ,

$$\pi(\theta) = \pi_{21}(\mu) \pi_{22}(\sigma^2) \pi_{23}(\lambda), \quad (3.20)$$

com a suposição de independência a priori. Escolhemos para $\pi_{21}(\mu)$ e $\pi_{22}(\sigma^2)$ as prioris usualmente consideradas no modelo probito normal. Por exemplo, $\mu \sim N(\mu_0, \tau_0^2)$ e $\sigma^2 \sim Inv-\chi^2(\omega, \kappa)$

(isto é, uma distribuição ji quadrado inversa) ou se a precisão definida como $1/\sigma^2$ é considerada, então uma distribuição gama inversa pode ser considerada.

Para o parâmetro de assimetria, como em Bazán *et al.* (2004a, 2004b), nós podemos considerar o modelo probito assimétrico generalizado em termos de $d_j = \frac{\lambda_j}{(1+\lambda_j^2)^{1/2}}$, $j = 1, \dots, I$. Esta parametrização, nós denotamos “*parametrização d*”, onde d_j assume valores no intervalo $[-1, 1]$. O fato do espaço paramétrico ser limitado permite especificar uma priori uniforme (não informativa) que é própria. Assim podemos considerar $d \sim U(-1, 1)$, ou equivalentemente, $\lambda \sim T(0, 0.5, 2)$, em que $T(a, b, \nu)$ denota a distribuição t-Student com locação a , escala b e ν graus de liberdade. (Use a transformação de variáveis $\lambda = \frac{d}{(1-d^2)^{1/2}}$). Quando a especificação do modelo é em termos de λ , denotamos por “*parametrização λ*”.

3.6.2 Esquema de dados aumentados para o amostrador de Gibbs (DAGS)

Usando a verossimilhança em (3.11) ou (3.12) e a distribuição a priori em (3.19) com (3.20) é possível obter a distribuição posterior e implementar um procedimento de estimação bayesiana que envolve a verossimilhança de Bernoulli. Contudo tal procedimento é complicado desde vez que as integrais envolvidas para obter a distribuição posterior marginal são difíceis de serem obtidas. Representações baseadas na aumentação de dados como considerado em Albert e Chib (1993) foram introduzidas na Seção 3.4. Considerando a estrutura latente envolvida, as distribuições condicionais completas a posteriori para o modelo probito assimétrico podem ser obtidas sem problemas para implementação de inferência Bayesiana via MCMC, usando a primeira função de verossimilhança aumentada dada em (3.14) na Seção 3.4. Contudo, algumas das condicionais completas não podem ser diretamente amostradas e algoritmos tais como o Metropolis-Hastings (Chib and Greenberg, 1995) precisam ser considerados. Um procedimento muito mais simples é considerar a função de verossimilhança aumentada alternativa dada (3.17) que considera variáveis latentes adicionais modificando a variável latente auxiliar Z_i , $i = 1, \dots, n$.

No restante da seção, nós desenvolvemos um procedimento computacional para o modelo probito assimétrico baseado nesta função de verossimilhança de dados aumentados alternativa. Duas parametrizações podem ser usadas. A especificação da verossimilhança completa para a *parametrização d* é dada como segue:

$$Z_i^* | x_i, y_i, \beta, \mu, \sigma^2, d \sim N(\mu + \eta_i - \sigma d v_i, (1 - d^2)\sigma^2) I(Z_i^*, y_i),$$

$$V_i \sim HN(0, 1),$$

$$\beta \sim \pi_1(\cdot),$$

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

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

e

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

A especificação de verossimilhança completa para a *parametrização* λ é dada como segue:

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

$$v_i \sim HN(0, 1),$$

$$\beta \sim \pi_i(\cdot),$$

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

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

e

$$\lambda \sim T(a, b, v).$$

As estruturas hierárquicas acima podem ser facilmente introduzidas no programa WinBugs. Note que todas as distribuições condicionais completas para o amostrador de Gibbs são fáceis de amostrar. Mas, em situações gerais μ e σ^2 devem ser especificados e λ é um parâmetro único de interesse associado com a escolha da ligação probito assimétrico generalizado. Este é o caso dos modelos *CDS* e *BBB pa*. Considerando este fato, a quarta e quinta linha na hierarquia são eliminadas. Também, note que quando $d_j = 0$ ou $\lambda_j = 0$, a estrutura hierárquica da verossimilhança aumentada que correspondente ao modelo probito segue eliminando a sexta linha na hierarquia. Se μ e σ^2 são desconhecidas e, portanto, devem ser estimados, modelos probitos assimétricos mais gerais podem ser considerados.

3.7 Uma aplicação

Nós ilustramos a abordagem Bayesiana desenvolvida neste capítulo usando dados da mortalidade de insetos apresentados em Bliss (1935). Estes dados referem-se ao número de insetos adultos mortos após 5 horas de exposição ao gás carbônico para várias concentrações deste gás. Estes dados bastante conhecidos na literatura foram também analisados por Prentice (1976), Czado (1994) e Stukel (1988) concluindo que suas ligação assimétrica são mais convenientes e melhoram significativamente as ligações logito e probito. Em Spiegelhalter *et al.* (1996) uma ligação cloglog é sugerida.

Nós consideramos as seguintes ligações: *CDS pa*, *BBB pa*, cloglog, logito e probito. Em todos os casos prioris não informativas para os parâmetros de regressão reparametrizados são usadas.

Isto é, como em Spiegelhalter *et al.* (1996) nós consideramos escalas diferentes para as priors no modelo dada por

$$\eta_i = \alpha^* + \beta(x_i - \bar{x}), \quad \text{com} \quad \alpha = \alpha^* - \beta\bar{x}, \quad \beta \sim N(0, 1000) \quad \text{e} \quad \alpha^* \sim N(0, 1000).$$

Entretanto nós consideramos para a parametrização *BBB pa*, as parametrizações d e λ , como especificado na Seção 3.5. Para a ligação *CDS pa*, nós consideramos só a parametrização λ com priori normal difusa para λ como dada em Chen *et al.* (1999).

Para todos as ligações, exceto para a cloglog e logito, a abordagem usando verossimilhança aumentada é considerada. Chen *et al.* (2000) sugerem que presença de autocorrelações entre valores da cadeia é esperada quando a verossimilhança aumentada é usada. Isto é significativo no modelo probito assimétrico como será também exibido no capítulo 5 quando aplicado o modelo *BBB pa* em a TRI. Considerando este fato, valores de salto (“thin”) acima de 400 são recomendados e conseqüentemente um número grande de iterações é necessário. Em todos os casos um tamanho efetivo da amostra igual a 2000 foi considerado. Número de iterações, valores de salto e tempo de execução são mostrados na Tabela 3.4.

A decisão sobre qual o valor de salto a ser usado foi baseada na condição de que a primeira autocorrelação para os parâmetros do modelo seja menor que 0.5. Vários critérios calculados usando o pacote CODA (Best *et al.* 1997), incluindo o critério proposto por Geweke (1992) foram usados para a avaliação da convergência.

3.7.1 Comparação de modelos

Para comparar os modelos, nós calculamos o Critério de Informação de Desvio (*DIC*), descrito em Spiegelhalter *et al.* (2002), o Critério de Informação Esperado de Akaike (*EAIC*) e o Critério de Informação Bayesiano (Schwarz) Esperado (*EBIC*), propostos em Brooks (2002) e a soma de quadrados dos residuais latentes (*SQRL*). Todos estes critérios são descritos no Apêndice B. O último critério não se aplica para a ligação cloglog e logito desde que residuais latentes não são disponíveis para estes modelo.

Os critérios mostrados na Tabela 3.1 confirmam os resultados de Collet (2003) p. 149, o qual nota, (usando deviança e residuais na abordagem freqüentista, disponíveis por exemplo em *Splus*) que o modelo log-log complementar ajusta melhor do que os modelos simétricos: logito e probito. Isto também pode ser verificado usando o Factor de Bayes aproximado como encontrado por Czado e Raftery (2001). Eles também encontram que a ligação cloglog é mais adequada que a ligação probito generalizada (Czado, 1994) e que tem um resultado similar do que a regressão logística usando um termo quadrático, modelos não implementados aqui. Não em tanto, da Tabela 3.1, parece clara a indicação que os modelos probito assimétricos são mais convenientes para os dados analisados pois obtiveram menores valores em todos os critérios do que a ligação

Tabela 3.1 Comparação dos dados sobre mortalidade dos Insetos data

modelos	Bur-in	DIC	EBIC	EAIC	SQRL
logito	374,40	376,4	386,8	378,4	
probito	373,40	375,4	385,8	377,4	482,2
cloglog	366,70	368,7	379,1	370,7	
CDS pa	160,40	273,4	178,9	166,4	481,4
BBB pa	292,3	257,5	310,8	298,3	251,1

cloglog. Note que *BBB pa* é melhor quando é considerado *DIC* e *SQRL* e que *CDS pa* é melhor quando considerados *EBIC* e *EAIC*. Nós consideramos que o *BBB pa* é mais adequado desde que *EBIC* e *EAIC* são penalizados com penalização dependente dos dados e *DIC* é penalizado pelo número efetivo de parâmetros (Spiegelhalter *et al.*, 2002) enquanto que *SQRL* mostra a discrepância entre as observações e as estimativas posteriores como é indicado no Apêndice A.

O uso da ligação probito assimétrica é mais conveniente para estes dados porque o modelo cloglog não tem um parâmetro que controla a taxa de crescimento (ou decréscimo) da probabilidade de sucesso (falha) da resposta binária e outros parâmetros associados com a locação e escala desta probabilidade. Além disso, estes modelos são facilmente implementados e não é necessário especificar intervalo de variação como os discutidos na introdução para as ligações assimétricas obtidas por modificação do preditor linear. Também, como provado na Seção 3.3, as distribuições a posteriori existem, o que pode não ser o caso de outras ligações assimétricas com mudança no preditor linear, como a devida a Stukel (Chen *et al.*, 1999).

3.7.2 Estimação de parâmetros

Na Tabela 3.2 apresentamos as estatísticas a posteriori dos parâmetros de regressão para os diferentes modelos considerados. Se nós olhamos somente os coeficientes estimados, então poderia ser natural concluir que os modelos considerados produzem estimativas diferentes. Mas uma comparação mais cuidadosa é comparar o “declive estimado” para os parâmetro (Green, 2003). Note que $(\frac{\partial E[y|\mathbf{x}]}{\partial x} = f(\eta)\beta)$ é uma medida para avaliar o efeito marginal dos parâmetros dado um mesmo valor do \mathbf{x} . Para obter uma estimativa desta medida nós consideramos o fator de escala $f(\bar{\eta})$, que corresponde à função de densidade avaliada nas médias a posteriori dos parâmetros a posteriori e da média das respostas. Considerando este fator o “declive” pode ser estimado ao se multiplicar por as estimativas baseadas nas médias a posteriori dos parâmetros. Considerando estas medidas para os parâmetros de regressão note que estas são similares entre os modelos.

Finalmente na Tabela 3.3, exibimos estatísticas a posteriori para os parâmetros de assimetria nos modelos probito assimétricos. Como era esperado, o valor do parâmetro de assimetria apresenta sinais diferentes para o *CDS pa* e *BBB pa*. Por outro lado, nós notamos também que

Tabela 3.2 *Estatísticas posteriores dos parâmetros de regressão para diferentes modelos para os dados de mortalidade de insetos*

modelos	parâmetros	média	mediana	95 % HPD		declive	$f(\bar{\eta})$
				Inferior	Superior		
logito	α	-60,94	-60,79	-71,52	-51,45	-7,41	0,12
	β	34,40	34,34	29,23	40,51	4,18	
probito	α	-35,13	-35,14	-39,95	-30,01	-7,50	0,21
	β	19,84	19,84	16,90	22,50	4,24	
cloglog	α	-39,66	-39,58	-45,65	-33,24	-7,45	0,19
	β	22,09	22,04	18,57	25,45	4,15	
CDS <i>pa</i>	α	-89,38	-85,95	-142,10	-42,55	-7,64	0,09
	β	48,62	46,79	23,97	77,27	4,16	
BBB <i>pa d</i>	α	-28,72	-28,83	-38,6	-19,62	-7,25	0,25
	β	16,00	16,09	10,57	21,67	4,04	

as parametrizações d e $lambda$ produzem resultados muito próximos, embora a parametrização $lambda$ é mais demorada quanto a convergência (veja a Tabela 3.4). Como exibido na Tabela 3.4, para os modelos probito assimétricos, um grande número de iterações são necessárias (em consequência um maior tempo de execução) e métodos como “over relax” (Neal, 1998) podem ser considerados para garantir a redução das autocorrelações. Este método gera amostras múltiplas em cada iteração e então escolhe uma que está negativamente correlacionada com o valor recente. O tempo de iteração pode ser incrementado, mas as autocorrelações dentro das cadeias são reduzidas e então poucas iterações resultam necessárias.

Tabela 3.3 *Estatística posterior do parâmetro de assimetria para os modelos probito assimétricos em dados de Mortalidade de insetos*

modelos	parâmetros	média	mediana	95 % HPD		Erro MC
				Inferior	Superior	
CDS <i>pa</i>	λ	4,19	4,02	1,507	6,792	0,07
BBB <i>pa d</i>	d	-0,48	-0,70	-0,998	0,314	0,02
	λ	-1,33	-0,99	-3,648	0,722	0,05
BBB <i>pa</i> λ	d	-0,48	-0,69	-0,995	0,482	0,02
	λ	-1,33	-0,99	-4,417	0,999	0,07

Tabela 3.4 *Desempenho de MCMC para diferentes modelos em dados de mortalidade de insetos*

modelos	salto	iterações	tempo (seg.)	tempo(seg.)	método regressão	método assimetria
			1000 it.	10000 it.		
logito	5	10000	9	89	GLM.Logit	
probito	5	10000	18	186	Slice.Std	
probito aug	5	10000	8	85	Normal.Std	
cloglog	5	10000	22	198	Slice.Std	
<i>CDS pa</i>	1050	2100400	16	162	Normal.Std	Normal.Std
<i>BBB pa d</i>	450	904000	23	220	Normal.Std	Slice.Interval
<i>BBB pa λ</i>	1200	2500400	29	302	Normal.Std	Metnormal.

Time in seconds to 10000 iterations an Pentium IV with 1800 MHZ and 256 Ram

3.8 Discussão Final

Neste capítulo é proposta uma nova ligação assimétrica para a resposta binária considerando a distribuição acumulada da distribuição normal assimétrica (Dalla-Valle, 2004). A ligação probito assimétrica generalizada introduzida tem como casos particulares a ligação probito assimétrica dada por Chen *et al.* (1999) (baseada na normal assimétrica *CDS*), a ligação probito assimétrica devida a Bazán *et al.* (2004a) (baseada na normal assimétrica *BBB*) e a ligação probito. A ligação probito assimétrica controla a taxa de crescimento (ou decréscimo) da probabilidade de sucesso (falha) da resposta binária considerando um parâmetro de assimetria da curva de resposta que é facilmente interpretado e outros parâmetros associados com a locação e escala desta probabilidade.

Nesta ligação, a probabilidade de sucesso é obtida considerando a fda de uma função de distribuição avaliada no preditor linear. O parâmetro de assimetria está associado com a distribuição escolhida e é independente do preditor linear, também, para a formulação do modelo não é necessária a especificação de uma estrutura latente linear.

Foram apresentados resultados sobre a existência das estimativas de máxima verossimilhança e das distribuições a posteriori dos parâmetros do modelo sob prioris uniformes impróprias. Estimação bayesiana foi desenvolvida e a implementação da metodologia foi facilmente obtida considerando as versão da verossimilhança aumentada proposta.

Como é estabelecido, para executar a inferência bayesiana na regressão binária, nós devemos adicionar a especificação de uma densidade a priori no vector de parâmetros de regressão β . Como Johnson e Albert (1999) mencionam, a especificação direta de uma densidade a priori para os parâmetros de regressão pode ser uma tarefa difícil devido ao efeito indireto que os parâmetros de regressão têm nas probabilidades de sucesso. Também, Paulino *et al.* (2003) mencionam que, numa análise bayesiana, a introdução de parâmetros adicionais levam a potencialmente

problemas sérios, porque estes parâmetros não estão relacionados diretamente com os dados como no caso em que escolhas competidoras da função de ligação devem ser consideradas, o que aconteceu na aplicação desenvolvida. Eles recomendam a elicitación de prioris por especialista e sugerem que métodos como os propostos por Bedrick *et al.* (1996) podem ser usados. Nesta classe de distribuições, prioris baseadas em opinião para a locação das probabilidades de sucesso p_i são avaliadas para casos particulares das covariáveis x_i , e esta informação é usada para construir uma priori para o vector de parâmetros de regressão β . Considerando a estimação das probabilidades de sucesso para vários modelos na Tabela 3.2, a elicitación de prioris para p_i por especialistas, pode ser importante para a escolha da função de ligação.

Aplicações em várias áreas em que as ligações simétricas não são justificadas podem ser obtidas com o modelo proposto como sera mostrado no capítulo 4 e 5 da tese em que se aplica a ligação para a Teoria da resposta ao item. Para outras aplicações em regressão veja Chen *et al.* (1999). Comparações de modelos simétricos e assimétricos são apresentadas usando os critérios *DIC*, *EAIC* e *EBIC* mas nós também introduzimos residuais latentes para os modelos e medidas de discrepância global como a soma de quadrados a posteriori de residuais latentes (*SQRL*) pode ser usado para comparação de modelos incluindo diferentes escolhas do modelo probito assimétrico como fora mostrado com o exemplo é dado.

Capítulo 4

A família normal assimétrica na teoria da resposta ao item

4.1 Introdução

A suposição de normalidade para as variáveis latentes e para a construção da curva característica do item (CCI) tem sido usada nos últimos 50 anos em muitos modelos psicométricos da Teoria da resposta ao item (TRI). Este capítulo introduz uma nova família de modelos assimétricos para a TRI, denominada família TRI normal assimétrica (TRI-NA). Esta família estende o modelo probito normal apresentado no Capítulo 2 de duas maneiras, considerando: a) CCI baseadas na ligação probito assimétrica generalizada apresentada no Capítulo 3, e b) que as variáveis latentes que modelam a habilidade individual tem distribuição normal assimétrica. Dependendo da existência ou não de assimetria em (a) ou (b), quatro modelos podem ser considerados, a saber:

- O modelo probito assimétrico-normal assimétrico (PANA),
- O modelo probito-normal assimétrico (PNA),
- O modelo probito assimétrico-normal (PAN),
- O modelo probito-normal (PN).

O primeiro é mais geral e inclui os demais como casos particulares.

Assim, a TRI-NA é uma família de modelos mais flexíveis para se ajustar a um conjunto de dados com respostas dicotômicas. A inferência Bayesiana é desenvolvida, usando-se técnicas de dados aumentados semelhante a introduzida nos Capítulos 2 e 3. Também, considera-se escolha de modelos utilizando os critérios *DIC*, *EAIC* e *EBIC*, e a soma de quadrados de residuais latentes *SQRL* apresentados no Apêndice B.

Como já foi dito no Capítulo 2, a TRI para itens dicotômicos modela um conjunto de respostas dicotômicas multivariadas obtidas a partir de n indivíduos avaliados num teste de k itens considerando uma variável latente comum para os indivíduos e um conjunto de parâmetros relativos aos itens do teste. A TRI modela a probabilidade de acerto das respostas obtidas em função de uma expressão não-linear (função de ligação) que liga a resposta binária com uma expressão linear entre a variável latente u e os parâmetros dos itens $\beta_j = (a_j, b_j)$.

Formalmente, seja y_{ij} a resposta dicotômica correspondente ao i -ésimo indivíduo, $i = 1, \dots, n$ no j -ésimo item j , $j = 1, \dots, k$ assumindo o valor 1, se a resposta é correta, e o valor 0 caso contrário. Assume-se que: $y_{ij} \sim \text{Bernoulli}(p_{ij})$ com $p_{ij} = P(y_{ij} = 1 \mid u_i, \beta_j)$ sendo a probabilidade de que o i -ésimo indivíduo seja capaz de responder o j -ésimo item corretamente. Como visto no Capítulo 2, na TRI considera-se que $p_{ij} = F(m_{ij})$, onde m_{ij} é uma função linear da variável latente dada por $m_{ij} = a_j u_i - b_j$, e, $F(\cdot)$, conhecida como CCI, é comum para todo i e todo j , e satisfaz a propriedade de *monotonicidade latente*.

No Capítulo 2, nós consideramos F como a acumulada da distribuição normal padrão que define a ligação probito. No entanto, como foi mostrado no capítulo 3, ligações assimétricas como o probito assimétrico podem ser adequadas para resposta binária.

CCI assimétricas foram propostas por Samejima (1997, 2000), a qual propõe a família de expoentes logísticos positivos que têm como caso particular o modelo logístico de dois parâmetros. Entretanto, nenhuma extensão assimétrica tem sido proposta para o modelo probito-normal de dois parâmetros.

Além disso, na TRI é assumido que $u_i \sim N(\mu, \sigma^2)$ o que estabelece que as variáveis latentes associadas com os indivíduos que fazem o teste é bem comportada isto é, suas habilidades são uma amostra aleatória desta distribuição (Albert, 1992). No entanto, como já foi dito na introdução, concordamos com Samejima (1997) de que esta suposição é muito restritiva para a modelagem da conduta humana.

Neste capítulo uma nova CCI é considerada utilizando a ligação probito assimétrica generalizada apresentada no capítulo 3. Além disso, assume-se que a variável latente tem distribuição normal assimétrica. Considerando-se simultaneamente ou separadamente, a CCI assimétrica ou as variáveis latentes assimétricas, constrói-se uma nova família de modelos TRI a qual denominamos família da Teoria da Resposta ao Ítem normal assimétrica (TRI-NA). O modelo probito-normal apresentado no Capítulo 2 segue como um caso especial.

O capítulo está organizado como segue. A Seção 4.2 introduz a família TRI-NA considerando o parâmetro de assimetria na CCI, chamado aqui parâmetro de penalidade, e considerando assimetria na variável latente. Na Seção 4.3 é discutido o ajuste de máxima verossimilhança para a família TRI-NA. Uma abordagem bayesiana é desenvolvida na Seção 4.4 usando a metodologia MCMC para simular das distribuições a posterior dos parâmetros de item e variáveis latentes. Duas metodologias serão consideradas.

4.2 Modelo TRI normal assimétrico

4.2.1 A CCI probito assimétrica

Uma nova CCI para modelos TRI é obtida considerando que a probabilidade condicional p_{ij} de uma resposta correta para o item j , dado o valor u_i da variável latente correspondente ao i -ésimo indivíduo, é igual a $p_{ij} = F_\theta(m_{ij})$, $i = 1, \dots, n$, F_θ é a fda da distribuição normal assimétrica

com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$, com μ parâmetro de localização, σ^2 parâmetro de escala e λ parâmetro de assimetria. F_θ^{-1} é denominada função de ligação.

Embora o uso da função de ligação proibito assimétrica generalizada seja bastante geral, pode ser conveniente, para não ter um modelo super parametrizado, assumir que $\mu = 0$ e $\sigma^2 = 1$. Neste caso $F_\theta(\cdot)$ corresponde à distribuição normal assimétrica padrão (Azzalini, 1985) e podemos escrever $F_\lambda(\cdot)$ que é a ligação *BBB pa*. Note que, deixar que o parâmetro de assimetria seja diferente para cada item é muito conveniente pois possibilita obter itens com CCI simétricos ou assimétricos. Assim, nós admitimos a possibilidade de contar com itens que têm CCI simétricos ou assimétricos. Considerando isto, nos temos que o parâmetro de assimetria é o único parâmetro associada a função de ligação que também é um parâmetro associado ao item, isto é, temos um novo parâmetro de item. Como será visto posteriormente, isto não modifica a interpretação dos tradicionais parâmetros de dificuldade e discriminação do item no modelo proibito-normal, mas incorpora um novo parâmetro de item que oferece interessantes interpretações dentro da TRI.

Assim, para a família de modelos TRI normais assimétricos é assumido que

$$p_{ij} = F_{\lambda_j}(m_{ij}), \quad i = 1, \dots, n \quad (4.1)$$

Na expressão (4.1), a probabilidade é expressa como função de u_i e os parâmetros parâmetros associados ao item j : $\beta_j = (a_j, b_j)$ e λ_j . Com o propósito de interpretar adequadamente a CCI proposta, considerando a propriedade A.1 no Apêndice A e o Capítulo 3, podemos escrever (4.1) como

$$F_\lambda(m_{ij}) = 2\Phi_2\left(\begin{pmatrix} m_{ij} \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -d_j \\ -d_j & 1 \end{pmatrix}\right), \quad (4.2)$$

em que $\Phi_2(\cdot)$ é a fda da normal bivariada e $|d_j| \leq 1$ é o coeficiente de correlação da normal bivariada. Segue diretamente da propriedade da normal assimétrica acumulada dada no Apêndice A que $d_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}$. Aqui, λ_j é o parâmetro de assimetria da normal assimétrica e d_j é uma parametrização (uma transformação 1-1) de λ_j , e portanto também é um parâmetro de assimetria. Nós utilizamos a seguinte notação simplificada $\Phi_2[(m_{ij}, 0)'; -d]$ para nos referir a distribuição normal bivariada em (4.2).

Observe que para o caso $\lambda_j = 0$ (normal bivariada com correlação 0) a expressão em (4.2) se reduz a $2\Phi(m_{ij})\Phi(0) = \Phi(m_{ij})$ obtendo-se assim o modelo proibito normal apresentado no Capítulo 2.

A Figura 4.1 apresenta algumas CCI proibito assimétricas quando fixamos os parâmetros $a_j = 1$ e $b_j = 0$ e variamos os valores do parâmetro de assimetria λ_j , $\lambda_j = -2, -1, 0, 1, 2$. Na Figura 4.2 apresentamos também algumas CCI proibito assimétricas considerando d no lugar de λ como parâmetro de assimetria, $d = -0.9, -0.7, -0.5, 0, 0.5, 0.7, 0.9$.

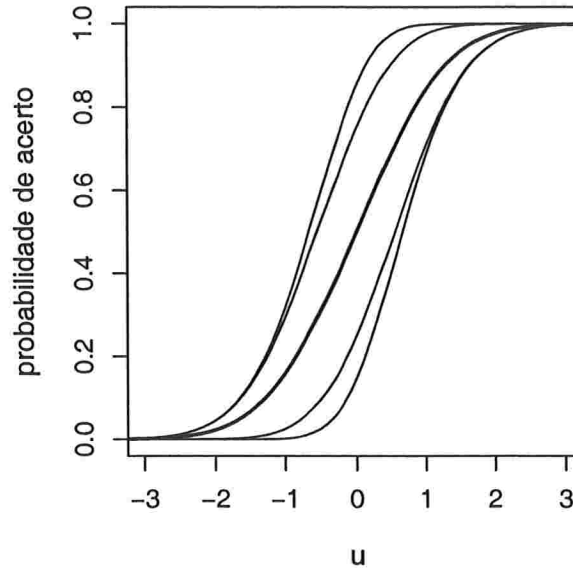


Figura 4.1 CCI probito assimétricos para $a = 1, b = 0$ e diferentes valores do parâmetro de assimetria $\lambda = -2, -1, 0, 1, 2$ (nesse ordem)

Note que em ambos casos para d ou $\lambda = 0$ a CCI é simétrica, e para d ou $\lambda > 0$ a CCI apresenta assimetria positiva, e para d ou $\lambda < 0$ apresenta assimetria negativa.

4.2.2 Distância entre as probabilidades de uma CCI probito assimétrica e CCI probito

A seguir analisamos a distância entre as probabilidades associadas com as CCI probito assimétrica p_{ij}^{pa} e probito p_{ij}^p para um valor dado da variável latente u_i e dos parâmetros dos itens a_j, b_j , isto é, $|p_{ij}^p - p_{ij}^{pa}|$.

Considerando a Propriedade 6 no Apêndice A, sabemos que esta distância é no máximo $\frac{1}{\pi} \arctan \lambda_j$. Isto significa que para um valor dado da variável latente, quando cresce a assimetria positiva, p_{ij}^{pa} decresce com respeito a p_{ij}^p . Caso contrário, se a assimetria negativa cresce, p_{ij}^{pa} é maior do que a p_{ij}^p .

Assim a CCI proposta é uma extensão natural da CCI do modelo TRI probito-normal pois introduz diretamente um parâmetro de assimetria para controlar a forma de crescimento das p_{ij} .

Como pode-se ver na Figura 4.1, as distâncias das probabilidades de resposta correta na

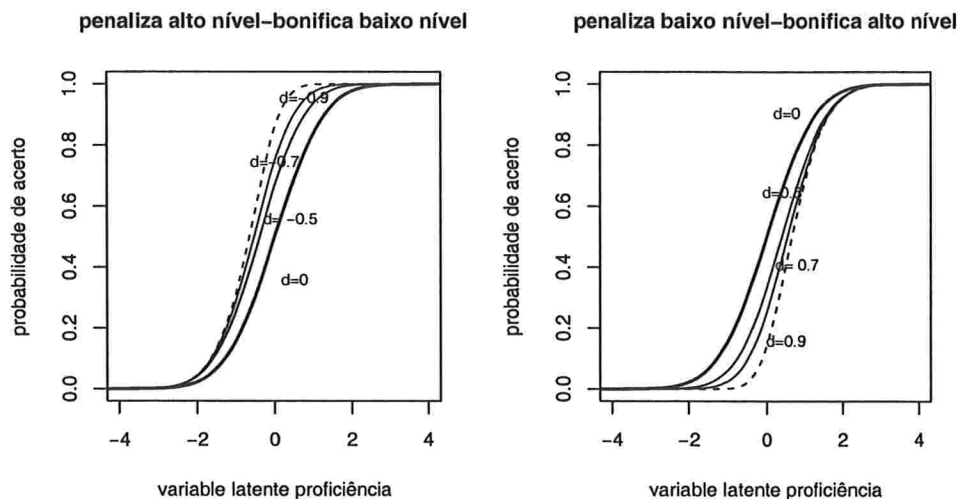


Figura 4.2 CCI probito assimétricos para $a = 1, b = 0$ e diferentes valores do parâmetro de assimetria d

CCI probito assimétrica e CCI probito variam também segundo os níveis da variável latente. Em presença de assimetria negativa, p_{ij}^{pa} e p_{ij}^p são mais similares para valores baixos da variável latente. Entretanto quando a variável latente assume valores altos, a distância das probabilidades se incrementa a favor do CCI probito assimétrico. Isto acontece porque a forma da S da CCI probito assimétrica cresce mais rapidamente do que a CCI probito. O contrário acontece em presença de assimetria positiva. Neste caso, p_{ij}^{pa} e p_{ij}^p são mais similares para valores altos da variável latente e apresentam maior distância para valores baixos a favor do CCI probito. Isto acontece porque a forma da S da CCI probito assimétrica cresce mais lentamente que a CCI probito.

Maiores comparações destas ligações no contexto da regressão binária são dadas no Apêndice C.

4.2.3 Interpretação dos parâmetros dos ítems na CCI probito assimétrica

Interpretação do parâmetro de assimetria

Um outro aspecto importante é analisar como se comportam os p_{ij} 's na CCI probito assimétrica com relação a pequenas mudanças na variável latente. Derivando p_{ij} em relação a u_i , temos a seguinte expressão

$$\frac{dp_{ij}}{du_i} = \frac{dF_{\lambda}(a_j u_i - b_j)}{du_i} = f_{\lambda}(a_j u_i - b_j) a_j = 2a_j \phi(a_j u_i - b_j) \Phi(\lambda_j(a_j u_i - b_j))$$

que corresponde à função de densidade da normal assimétrica para U_i com locação $\frac{b_j}{a_j}$, escala $\frac{1}{a_j}$ e parâmetro de assimetria λ_j . Note que para valores fixos dos parâmetros dos itens a_j e b_j , as mudanças de p_{ij} dependem do sinal de λ e do valor da variável latente. Do fato da CCI proibito assimétrica ser obtida considerando uma fda temos que a CCI é uma função monótona crescente da quantidade U_i a qual é assumida unidimensional. Isto significa que os modelos TRI-NA são modelos de variáveis latentes monótonos unidimensionais (Junker e Ellis, 1997).

Da Figura 4.1 e considerando um valor da variável latente igual a zero, note que quando a assimetria é mais negativa, a probabilidade de resposta correta p_{ij} é maior do que 0.5, e quando a assimetria é mais positiva p_{ij} é menor do que 0.5.

Em presença de assimetria negativa, a medida que a habilidade latente cresce, a mudança da probabilidade de acerto como consequência de uma mudança da habilidade vai diminuindo. Isto pode ser interpretado no contexto educacional da seguinte maneira. Quando existe assimetria negativa, uma mudança de uma unidade na medida de habilidade se traduz num incremento maior da probabilidade de acerto para os estudantes com menor habilidade do que para os estudantes com maior habilidade. Por outro lado, em presença de assimetria positiva, a medida que a habilidade latente cresce, a mudança de probabilidade de acerto como consequência de uma mudança da habilidade vai aumentando.

Considerando este comportamento, o parâmetro de assimetria pode ser psicometricamente interpretado como um parâmetro de penalidade (bonificação) da probabilidade de resposta correta. Assim, um item com parâmetro de assimetria negativo penaliza (bonifica) estudantes com grandes (pequenos) níveis da variável latente e um item com assimetria positiva bonifica (penaliza) indivíduos com grandes (pequenos) níveis da variável latente (veja-se Figura 4.2). Portanto, nós denominamos o parâmetro de assimetria genericamente de *parâmetro de penalidade*.

Interpretação dos parâmetros a e b

A Figura 4.3 apresenta seis diferentes CCI assumindo diferentes valores para λ_j . As primeiras três curvas consideram variações adicionais no parâmetro b_j com a_j fixo. As últimas três CCI consideram variações nos parâmetros a_j com b_j fixo. O parâmetro b_j é chamado o intercepto do item ou parâmetro de dificuldade. Este parâmetro controla os níveis de dificuldade do item. Se nós fixamos os parâmetros a_j e λ_j e incrementamos o valor de b_j , a forma básica da CCI não muda mas é trasladada para a direita. As curvas C1, C2 e C3 na Figura 4.3 correspondem, respectivamente, a $(a_j = 1, b_j = -1)$, $(a_j = 1, b_j = 0)$ e $(a_j = 1, b_j = 1)$. Um item de um teste com altos valores negativos de b_j (curva C1) corresponde a um item fácil no qual os indivíduos com pequenas médias na variável latente apresentam relativamente baixa probabilidade de resposta correta. Por outro lado, um item com um valor grande de b_j (curva C3) é difícil, devido ao fato de que indivíduos com grandes níveis da variável latente apresentam, relativamente, baixa probabilidade de resposta correta. A conduta deste parâmetro para a ligação proibito assimétrica pode ser observada na Figura 4.3 para diferentes valores do parâmetro de assimetria.

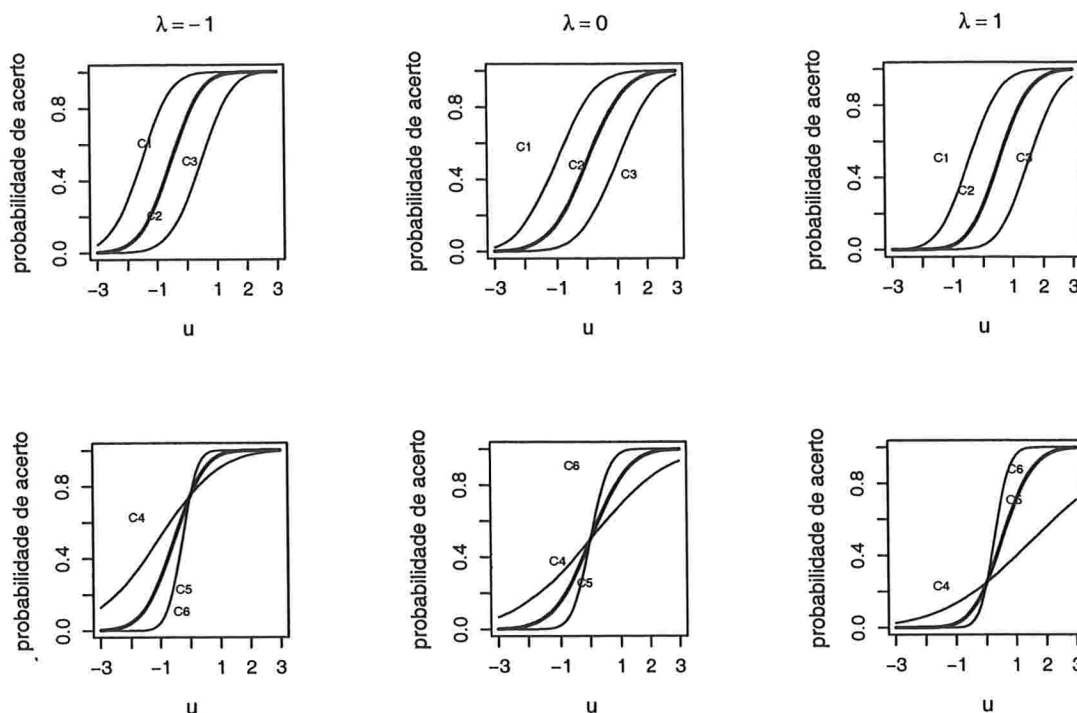


Figura 4.3 CCI probito assimétricos para diferentes valores dos parâmetros b e a considerando assimetria

O parâmetro a_j é chamado *parâmetro de inclinação ou de discriminação*. Este parâmetro controla a inclinação da CCI. Se consideramos uma CCI com valores de b_j e λ_j fixos, a curva é mais empinada para valores crescentes de a_j . As curvas C4, C5, e C6 na Figura 4.3 correspondem respectivamente a $(a_j = 0.5, b_j = 0)$, $(a_j = 0, b_j = 0)$ e $(a_j = 1, b_j = 0)$. Uma CCI mais empinada corresponde a um item que é altamente discriminante entre estudantes de menor e maior nível da variável latente. A probabilidade de acerto muda rapidamente quando maior é o valor da variável latente num intervalo em torno de 0. Isto significa que a probabilidade de resposta correta muda apenas uma pequena quantidade quando a variável latente vai de um estudante com baixo valor da variável latente a um estudante com alto valor da variável latente. Um item com um pequeno valor de a_j é relativamente de pobre discriminação entre estudantes para mudanças na variável latente.

Em resumo, a interpretação dos parâmetros β é a mesma para CCI probito assimétrica e para CCI probito no modelo TRI-NA. Seguindo diferentes propostas na literatura, nós reparametrizamos os parâmetros do item considerando $a = \alpha$ e $b = \alpha\beta$ tal que $m_{ij} = a_j u_i - b_j$ com $\beta_j = (a_j, b_j)$ os clássicos parâmetros dos item correspondentes ao j -ésimo item. De acordo a Cook *et al.* (2002)

e Baker (1992) esta parametrização resulta em cálculos mais estáveis. Nos usamos a notação $\mathbf{a} = (a_1, \dots, a_n)^T$ e $\mathbf{b} = (b_1, \dots, b_n)^T$.

4.2.4 Variável latente distribuída assimetricamente

Uma nova distribuição da variável latente U_i correspondente ao i -ésimo indivíduo pode ser definida considerando

$$U_i \sim NA(\theta_u), \quad i = 1, \dots, n \quad (4.3)$$

onde $\theta_u = (\mu_0, \sigma_0^2, \kappa)$ denota a distribuição normal assimétrica com parâmetro de locação $-\infty < \mu_0 < \infty$, parâmetro de escala $\sigma_0^2 > 0$ e parâmetro de assimetria $-\infty < \kappa < \infty$. No apêndice A apresentamos algumas propriedades desta distribuição. Aqui sua função de densidade é denotada por $g_{\theta_u}(u)$. Assim, nós consideramos que a variável latente segue uma distribuição normal assimétrica com parâmetro de assimetria comum para os indivíduos. Isto é, nós opinamos que assimetria não muda de um indivíduo a outro desde que é uma propriedade da distribuição da variável latente para a população completa da qual o indivíduo é extraído.

A Figura 4.4 mostra as funções de densidade das variáveis latentes para diferentes valores dos hiperparâmetros μ_0 e σ_0 e para o parâmetro de assimetria κ . As três curvas no lado direito são exemplos de parâmetros de assimetria κ positivos, que modelam variáveis latentes concentradas em valores baixos. As três curvas no lado esquerdo são exemplos de parâmetros de assimetria κ negativos que modelam variáveis latentes concentradas em valores altos. Curva $N(0,1)$ é apresentada como uma referência em todos os casos. Note que as curvas mostradas na Figura 4.4 podem corresponder à distribuição da variável latente correspondente à conduta humana em diferentes casos, como observada em Micceri (1989). Exemplos de tais condutas são depressão (veja-se Riddle *et al.* 2002) e ansiedade (veja-se Zaidler *et al.*, 2003) em que alguma assimetria é esperada considerando uma população não clínica. Além disso, em vários contextos educacionais variáveis preditoras relativas à proficiência das escolas podem ser assimetricamente distribuídas, como notado em Hashimoto (2002) e Vianna (2003). Portanto, a distribuição normal assimétrica é um modelo flexível para modelar variáveis latentes e além disso acomoda à distribuição normal como caso especial.

4.2.5 A família TRI normal assimétrica

A família TRI-NA é definida considerando que

$$y_{ij}|u, a_j, b_j, \lambda_j \sim \text{Bern}(p_{ij}) \quad (4.4)$$

com p_{ij} definida em (4.1), $i = 1, \dots, n$ and $j = 1, \dots, k$, e satisfazendo os princípios de *independência condicional latente* que considera que para o i -ésimo indivíduo, y_{ij} são condicionalmente independentes dado u_i e a *independência entre as respostas dos indivíduos* (veja-se Capítulo 2).

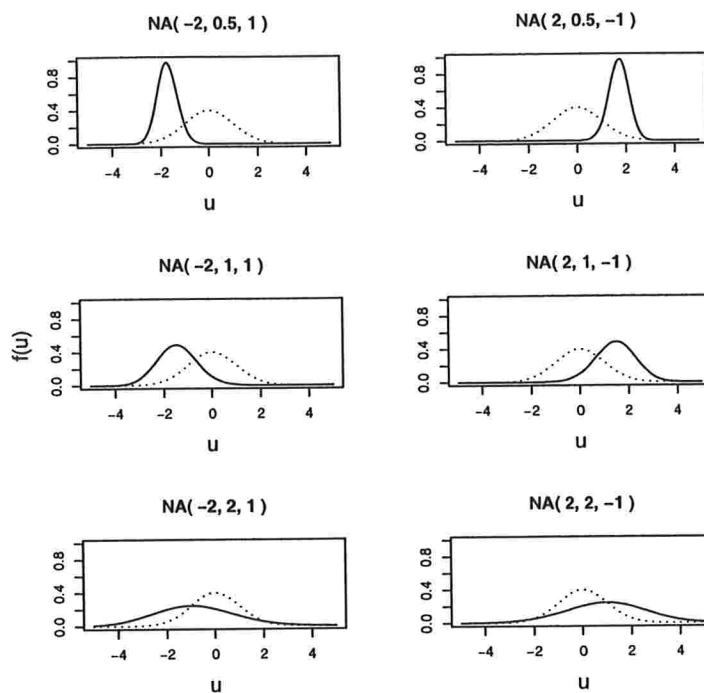


Figura 4.4 função de densidade da normal assimétrica $\theta = (\mu, \sigma^2, \lambda)$

É considerado que as variáveis latentes são igualmente distribuídas como em (4.3). Além disso, como a família TRI-NA é definida considerando a especificação de uma nova CCI assimétrica e distribuições assimétricas para a variável latente, quatro modelos TRI-NA podem ser especificados: a) o modelo probito assimétrico-normal assimétrico (PANA), b) o modelo probito assimétrico-normal (PAN) com $\lambda_j = 0$, c) o modelo probito-assimétrico normal (PNA) com $d_j = 0$; d) o modelo probito-normal (PN) no qual $\lambda_j = 0$ e $d_j = 0$. Um aspecto interessante dos modelos formulados acima é a flexibilidade na detecção de itens especificados em concordância para CCIs com ligações probito assimétricas e itens especificados de acordo a CCIs com ligações probito. Assim, a família de modelos TRI-NA apresentam modelos mais flexíveis para o modelagem matemática da conduta psicológica e educacional, baseada em *processos dedutivos* (Samejima, 1997) e completa a pesquisa para modelos que ajustam a conduta em questão teoricamente. *Processos indutivos* (Samejima, 1997) como TRI não paramétrico são tão importantes quanto os processos dedutivos e usos dinâmicos de ambos processos nós permitem simular a conduta humana.

4.3 Estimação por máxima verossimilhança

A seguir, utilizaremos a notação $\mathbf{a} = (a_1, \dots, a_k)^T$, $\mathbf{b} = (b_1, \dots, b_k)^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$, $\mathbf{y} = (y_{11}, \dots, y_{kn})^T$. Então, a função de verossimilhança para a família TRI-NA é dada por

$$L(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}) = \prod_{i=1}^n \prod_{j=1}^k F_{\lambda_j}(m_{ij})^{y_{ij}} (1 - F_{\lambda_j}(m_{ij}))^{1-y_{ij}} \quad (4.5)$$

com F_{λ_j} dado na expressão 4.2.

Como no caso do modelo PN, o cálculo das estimativas de máxima verossimilhança e da matriz de informação de Fisher é simplificada se nós dispormos de procedimentos para o cálculo da distribuição normal bivariada. Estimativas de máxima verossimilhança conjunta (MVC) similares as usadas com o modelo PN podem ser implementadas para a estimação de \mathbf{u} , $\boldsymbol{\beta}$ e $\boldsymbol{\lambda}$ (ou d). Uma alternativa é obter estimativas iterativamente, maximizando a verossimilhança sobre os parâmetros do item supondo fixas as variáveis latentes e vice-versa (Baker, 1992) usando o valor obtido na primeira maximização. Entretanto, as limitações deste método são bastante conhecidas (veja-se Baker, 1992, e Bock e Aitkin, 1981). Assumindo conhecidos a variável latente nós podemos implementar estimativas de máxima verossimilhança separada para os parâmetros dos itens, e viceversa, assumindo conhecidos os parâmetros dos itens nós podemos implementar estimativas de máxima verossimilhança separada para a variável. Um outro procedimento, baseado na estratégia de “divide-e-conquistarás” (Patz e Junker, 1999), já comentada no Capítulo 2, é estimar só os parâmetros do item e $\boldsymbol{\lambda}$ (ou d) é o método de máxima verossimilhança marginal (MVM). O procedimento pode ser desenvolvido implementando um algoritmo tipo EM como considerado em Bock e Aitkin (1981) tratando as variáveis latentes como dados perdidos e requerendo a especificação de uma distribuição para a variável latente U . Como nós propomos considerar uma distribuição normal assimétrica com parâmetro de assimetria comum para os indivíduos, i.e. $U_i \sim NA(\theta_u)$, $i = 1, \dots, n$ segue que a função de verossimilhança marginal para a família TRI-NA com interesse nos parâmetros pode ser escrita como:

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}, \mathbf{u}) = \int_{\mathbb{R}^n} \prod_{i=1}^n \prod_{j=1}^k F_{\lambda_j}(m_{ij})^{y_{ij}} (1 - F_{\lambda_j}(m_{ij}))^{1-y_{ij}} g_{\theta_u}(u_i) du_i \quad (4.6)$$

Procedimentos de estimação baseados no algoritmo EM podem ser implementados a través do uso de quadratura gaussiana como uma maneira de aproximar a integral múltipla na verossimilhança (4.6). O mesmo pode ser feito para a verossimilhança expressa em termos da distribuição normal bivariada truncada com correlation d_j . Como mencionado em Patz e Junker (1999) quando aumenta a complexidade do modelo, como no caso da família TRI-NA, ou o uso de algoritmos EM tornam-se menos satisfatórios e é difícil incorporar a incerteza (erros padrão) na estimação dos parâmetros de ítem e no cálculo da incerteza (erros padrão) para a inferência dos examinados (indivíduos), e por causa desta incerteza não existe um modo de avaliar a extensão dos erros padrão para que a inferência dos examinados seja mais precisa.

No Capítulo 3 discutimos as condições para a existência de estimativas de máxima verossimilhança quando usado ligações proibito assimétricas, estas condições podem ser estendidas naturalmente para o modelo TRI-NA.

Considerando a Propriedade A.2 no Apêndice A, nós temos que para $\theta = (0, 1, \lambda)$

$$F_{\lambda_j}(m_{ij}) = \int_0^{\infty} \Phi\left(m_{ij}\sqrt{1+\lambda_j^2} - \lambda_j v_{ij}\right) g(v_{ij}) dv_{ij} \quad (4.7)$$

em que $\Phi(\cdot)$ denota a fda da distribuição normal padrão e $g(v) = \frac{2}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} I(v > 0)$ é a fda da distribuição Normal positiva.

Conseqüentemente, a função de verossimilhança para a família TRI-NA pode ser escrita como

$$L(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}, \mathbf{v}) = \prod_{i=1}^n \prod_{j=1}^k \int_0^{\infty} \Phi\left(m_{ij}\sqrt{1+\lambda_j^2} - \lambda_j v_{ij}\right)^{y_i} \left[1 - \Phi\left(m_{ij}\sqrt{1+\lambda_j^2} - \lambda_j v_{ij}\right)\right]^{1-y_{ij}} g(v_{ij}) dv_{ij}. \quad (4.8)$$

Uma versão aumentada da verossimilhança é obtida, considerando-se as variáveis auxiliares v_{ij} , por

$$L(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}, \mathbf{v}) = \prod_{i=1}^n \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1+\lambda^2} - \lambda v_i\right)^{y_i} \left[1 - \Phi\left(\frac{\eta_i - \mu}{\sigma} \sqrt{1+\lambda^2} - \lambda v_i\right)\right]^{1-y_i} g(v_i) dv_i. \quad (4.9)$$

4.4 Estimação Bayesiana

4.4.1 Especificação de Prioris

Para implementar o procedimento de estimação bayesiana nós temos que incorporar distribuições a priori para $\mathbf{u}, \boldsymbol{\beta}$ e $\boldsymbol{\lambda}$. Nós consideramos a seguinte classe general de distribuições a priori independentes:

$$\pi(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \pi_1(\mathbf{u})\pi_2(\boldsymbol{\beta})\pi_3(\boldsymbol{\lambda}). \quad (4.10)$$

Além disso, seguindo propostas usualmente consideradas na literatura, assumimos independência a priori entre as habilidades dos indivíduos e entre os ítems. Assim temos:

$$\pi(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^k g_{2j}(\beta_j) g_{3j}(\lambda_j) \quad (4.11)$$

Adicionalmente supomos que $g_{2j}(\beta_j) = g_{21j}(a_j)g_{22j}(b_j)$ em que $g_{21j}(\cdot)$ e $g_{22j}(\cdot)$ devem ser próprias a fim de garantir que a distribuição posterior conjunta seja própria (Albert e Ghosh, 1999, pg. 178), Ghosh *et al* (2001). Seguindo as propostas usuais consideradas na literatura (ver Rupp *et al.* 2004), nós consideramos $g_{21j}(\cdot)$ como a densidade da $N(\mu_a, \sigma_a^2)I(a_j > 0), j = 1, \dots, k$ e $g_{22j}(\cdot)$ como a densidade da $N(0, \sigma_b^2), j = 1, \dots, k$ e, portanto, $g_{2j}(\cdot)$ é a densidade da $N_2(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ com vetor de médias $\boldsymbol{\mu}_\beta = (\mu_a, 0)$ e matriz de covariâncias $\boldsymbol{\Sigma}_\beta = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$. Para maiores detalhes acerca das prioris para os parâmetros a_j e b_j revise o Capítulo 2.

Nós consideramos $g_{1i}(\cdot)$ como a densidade de $NA(\theta)$, $i = 1, \dots, n$ com $\theta = (0, 1, \kappa)$ e $g_3 \equiv \phi_{SN}(\lambda; \omega)$ com $\mu_a, s_a, s_b, \kappa, \omega$ valores conhecidos. No entanto, para estimar κ e ω , uma estrutura hierárquica precisa ser introduzida considerando distribuições para esses hiperparâmetros.

Usando a verossimilhança tipo Bernoulli dada em (4.5) e as distribuições a priori especificadas acima poderia-se implementar um procedimento de estimação bayesiana usando ARS semelhante ao procedimento para o modelo PN. Entretanto, tal esquema não é possível em WinBUGS, porque, as integrais envolvidas para obter a distribuição posterior marginal envolvem a função acumulada da normal assimétrica ou da normal bivariada, as quais não são disponíveis atualmente nesse programa. Nós usamos um esquema baseado em dados aumentados, como considerado em Albert (1992) e Sahu (2002). Esta metodologia permite a implementação do MCMC de uma maneira simples que simplifica eficientemente o amostragem das distribuições a posterior marginais.

4.4.2 Algoritmo de dados aumentados

Como no caso do modelo de regressão binária, quando uma ligação probito assimétrica generalizada é usada, nós podemos obter duas novas versões da função de verossimilhança baseado em aumento de dados. Nosso algoritmo também é similar ao algoritmo apresentado no capítulo 2 para o modelo PN e é obtido usando a representação estocástica devida a Henze (1986).

Proposição 6. *O modelo TRI probito assimétrico, envolvendo k itens e n indivíduos, com $y_{ij} \sim \text{Ber}(p_{ij})$ e $p_{ij} = \Phi(m_{ij}, \lambda_j)$ na qual $m_{ij} = a_j u_i - b_j$, é equivalentemente definido considerando que*

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0, \end{cases} \quad (4.12)$$

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

Prova: A prova usa o fato de que $1 - F_{-\lambda}(z) = F_{\lambda}(-z)$ (veja-se propriedade E em Azzalini, 1985) e é similar à dada em Albert (1992) e no Capítulo 2 para o modelo probito-normal. \square

Claramente, no caso especial $\lambda_j = 0$, $j = 1, \dots, k$, segue o correspondente resultado do modelo PN dado no Capítulo 2. A variável latente Z_{ij} é introduzida para evitar trabalhar com a verossimilhança tipo Bernoulli. Entretanto, note que o parâmetro de assimetria com a variável latente auxiliar é o oposto do parâmetro de assimetria da CCI. A seguir, nós usamos a notação $\mathbf{Z} = (Z_{11}, \dots, Z_{kn})^T$. A função de verossimilhança aumentada para o modelo TRI probito assimétrico com $\mathbf{D} = (\mathbf{Z}^T, \mathbf{y}^T)^T$, \mathbf{a} e \mathbf{b} é dada por

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda | \mathbf{D}) = \prod_{i=1}^n \prod_{j=1}^k f_{\theta}(Z_{ij}) I(Z_{ij}, y_{ij}), \quad (4.13)$$

em que $\theta = (m_{ij}, 1, -\lambda_j)$ e $m_{ij} = a_j u_i - b_j$, e $I(Z_{ij}, y_{ij}) = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq$

$0)I(y_{ij} = 0)$, $j = 1, \dots, k$ e $i = 1, \dots, n$ corresponde à densidade condicional $p(y_{ij}|Z_{ij})$ usada para a derivação de (4.13). Note que, se $\lambda_j = 0$ então a função de verossimilhança (4.13) é similar à do modelo probito-normal apresentada no Capítulo 2.

Uma maneira alternativa de escrever o modelo TRI probito assimétrico é apresentada a seguir.

Proposição 7. *O modelo TRI probito normal assimétrico, com k itens e n indivíduos, com $y_{ij} \sim \text{Ber}(p_{ij})$ e $p_{ij} = \Phi_{NA}(m_{ij}, \lambda_j)$ na qual $m_{ij} = a_j u_i - b_j$, é equivalente a considerar que*

$$y_{ij} = \begin{cases} 1, & Z_{ij}^* > 0; \\ 0, & Z_{ij}^* \leq 0, \end{cases},$$

em que $Z_{ij}^* \sim N(-\delta_j V_{ij} + m_{ij}, 1 - \delta_j^2)$, e $V_{ij} \sim HN(0, 1)$, $j = 1, \dots, k$ e $i = 1, \dots, n$.

Prova: Definindo $Z_{ij}^* \doteq Z_{ij}|V_{i,j}$ a representação estocástica para a distribuição normal assimétrica (Henze, 1986) para Z_{ij} , em que $Z_{ij} \sim NA(m_{ij}, 1, -\lambda_j)$, $j = 1, \dots, k$, $i = 1, \dots, n$, como a dada na Proposição 6. Então, a prova segue pelo uso do Observação 1 na proposição A.1 para a distribuição normal assimétrica apresentada no Apêndice A. \square

Nós consideramos agora a função de verossimilhança aumentada envolvendo variáveis latentes $\mathbf{Z}^* = (Z_{11}^*, \dots, Z_{kn}^*)^T$. Seja $\mathbf{d} = (d_1, \dots, d_k)^T$. A função de verossimilhança de $(\mathbf{u}^T, \mathbf{a}^T, \mathbf{b}^T, \mathbf{d}^T)^T$ com dados aumentados $D = (\mathbf{Z}^*, \mathbf{X}, \mathbf{y})$ é dada por

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b}, \mathbf{d}|D) = \prod_{i=1}^n \prod_{j=1}^k \phi(Z_{ij}^*; -d_j V_{ij} + m_{ij}, 1 - d_j^2) I(Z_{ij}^*, y_{ij}) \phi(V_{i,j}; 0, 1) I(V_{i,j} > 0) \quad (4.14)$$

na qual $m_{ij} = a_j u_i - b_j$ e $I(c, d)$ é igual a 1 quando $\{c > 0, d = 1\}$ ou $\{c \leq 0, d = 0\}$ e é zero caso contrário.

4.4.3 Distribuições condicionais completas

As versões de verossimilhança aumentadas na Seção 4.4.2 e as distribuições a priori especificadas na Seção 4.4.1 levam a expressões das distribuições condicionais completas que podem ser usadas para o uso do amostrador de Gibbs. No entanto elas podem ser mais o menos complexas de implementar.

Proposição 8. Considerando a função de verossimilhança (4.13) e as distribuições a priori dadas na Seção 4.4.1. Na família TRI-NA, as distribuições condicionais completas para o amostrador de Gibbs são dados por:

- $\pi(Z_{ij}|U_i, \beta_j, \lambda_j, D_{obs}) \propto \phi_{NA}(Z_{ij}; m_{ij}, 1, -\lambda_j) I(Z_{ij}, y_{ij})$
- $\pi(U_i|Z_i, \beta, \lambda, D_{obs}) \propto \phi(U_i; m_{U_i}, v_{U_i}) \psi(U_i)$

em que: $\psi(U_i) = \prod_{j=1}^I \Phi(-\lambda_j Z_{ij} + b_j - U_i a_j) \Phi(\kappa U_i)$

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

- $\pi(\beta_j | \mathbf{u}, \mathbf{Z}_j, \lambda_j, D_{obs}) \propto \phi_2(\beta_j; \mathbf{m}_{\beta_j}, \mathbf{v}_{\beta_j}) \psi(\beta_j)$

onde: $\psi(\beta_j) = \prod_{i=1}^n \Phi(-\lambda_j Z_{ij} - \mathbf{W}'_i \beta_j)$

com $\mathbf{m}_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1} [\mathbf{W}'\mathbf{Z}_j + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta]$, $\mathbf{v}_{\beta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1}$, em que $\boldsymbol{\mu}_\beta =$

$\begin{pmatrix} \mu_a \\ 0 \end{pmatrix}$, $\Sigma_\eta = \begin{bmatrix} S_a^2 & 0 \\ 0 & S_b^2 \end{bmatrix}$ e $\mathbf{W} = (\mathbf{u}, -1)$ com $\mathbf{W}'_i = (U_i, -1)$

- $\pi(\lambda_j | \mathbf{u}, \mathbf{Z}_j, \beta_j, D_{obs}) \propto \phi(\lambda_j) \psi(\lambda_j)$ em que

$\psi(\lambda_j) = \prod_{i=1}^n \Phi(-\lambda_j Z_{ij} - m_{ij}) \Phi(\omega \lambda_j)$

Prova: De (4.13) e das prioris especificadas na Seção 4.1.1, temos que a distribuição conjunta a posteriori é dada por:

$$f(\mathbf{u}, \boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^I \phi_{NA}(Z_{ij}; m_{ij}, 1, -\lambda_j) I(Z_{ij}, y_{ij}) \\ \prod_{j=1}^I \phi_{NA}(U_i; \kappa) \prod_{i=1}^n \phi_2(\beta_j; \boldsymbol{\mu}_\beta, \Sigma_\beta) \phi_{SN}(\lambda_j; \omega)$$

A prova segue substituindo as densidades $\phi_{SN}(\cdot)$ pela expressão $2\phi(\cdot)\Phi(\cdot)$ e identificando na distribuição conjunta os termos que contêm os parâmetros desejados na condicional e agrupando as densidades correspondentes $\phi(\cdot)$ e as distribuições acumuladas $\Phi(\cdot)$ separadamente. No caso de agrupar as $\phi(\cdot)$ temos o resultado dado em Johnson e Albert (1999) para o caso probito-normal nas primeiras três condicionais, e no caso de agrupar as $\Phi(\cdot)$ temos uma função $\psi(\cdot)$ que depende do parâmetro mas é uniformemente limitada (as distribuições acumuladas são uniformemente limitadas). □

Neste esquema, as distribuições condicionais obtidas não correspondem a um procedimento direto do amostrador de Gibbs. A primeira condicional corresponde a distribuição normal assimétrica truncada que pode ser simulada facilmente usando a biblioteca *sn* dada por Azzalini (2003) para o programa *R* ou *Matlab*, entretanto, as outras distribuições condicionais não têm forma fechada, e precisa-se de um outro algoritmo para simular delas como, por exemplo, Metrópolis-Hasting. Estas condicionais satisfazem o requerimento de Chib e Greenberg (1995) correspondente à terceira família de densidades geradas candidatas para implementar o algoritmo Metrópolis-Hasting. Esta família usa o conhecimento da distribuição condicional digamos $\pi(t)$ que é da forma $\pi(t) \propto \psi(t)h(t)$ em que $h(t)$ é uma densidade que pode ser amostrada facilmente e $\psi(t)$ é uniformemente limitada. Neste caso, a probabilidade de mover-se na cadeia é dada por $\alpha(x, y) = \min\{\frac{\psi(y)}{\psi(x)}, 1\}$ em que y é gerado da $h(t)$ proposta e x é o antigo valor. No modelo, as condicionais de U_i, β_j e λ_j são da forma $\phi(\cdot)\psi(\cdot)$ com $\phi(\cdot)$ densidade da normal

univariada no caso das condicionais de U_i e de λ_j e da normal bivariada no caso da condicional de η_j que são fáceis de simular.

Outro aspecto importante dos resultados da Proposição 8, é que recuperamos o modelo proibit-normal quando é assumido ausência de assimetria para a variável latente U_i ($\kappa = 0$) e para a função de ligação F_{ij} ($\lambda_j = 0$). Nosso procedimento generaliza os procedimentos implementados em MCMCpack para R (Martin e Quinn, 2003) e Matlab (Johnson e Albert, 1999) para o modelo proibito normal desenvolvido no Capítulo 2.

Proposição 9. Considerando a função de verossimilhança (4.13) e as distribuições a priori dadas na Seção 4.4.1 para \mathbf{u} e β , e as distribuições a priori uniforme e independentes para \mathbf{d} no modelo TRI normal assimétrico, as distribuições condicionais completas para o amostrador de Gibbs são dadas por:

- $\pi(Z_{ij}^*|U_i, \beta_j, d_j, V_{ij}, D_{obs}) \propto \phi(-d_j V_{ij} + m_{ij}, 1 - d_j^2) I(Z_{ij}^*, y_{ij})$
- $\pi(V_{ij}|U_i, \eta_j, d_j, Z_{ij}^*, D_{obs}) \propto \phi(d_j(m_{ij} - Z_{ij}^*), 1 - d_j^2) I(V_{ij} > 0)$
- $\pi(U_i|Z_j^*, V_j, \beta, \mathbf{d}, D_{obs}) \propto \phi(U_i; m_{U_i}, v_{U_i}) \Phi(\kappa U_i)$
em que: $m_{U_i} = \frac{\sum_{i=1}^I a_j (Z_{ij}^* + d_j V_{ij} + b_j)}{1 + \sum_{i=1}^I a_j^2 - d_j^2}$, $v_{U_i} = \frac{1 - d_j^2}{1 + \sum_{i=1}^I a_j^2 - d_j^2}$
- $\pi(\beta_j|\mathbf{u}, Z_j^*, V_j, d_j, D_{obs}) \propto \phi_2(\eta_j; \mathbf{m}_{\eta_j}, v_{\eta_j})$
com
 $\mathbf{m}_{\eta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1} [\mathbf{W}'(Z_j^* + d_j V_j) + \Sigma_\eta^{-1} \mu_\eta]$,
 $v_{\eta_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\beta^{-1}]^{-1}$, em que $\mu_\beta = \begin{pmatrix} \mu_a \\ 0 \end{pmatrix}$, $\Sigma_\beta = \begin{bmatrix} S_a^2 & 0 \\ 0 & S_b^2 \end{bmatrix}$ e $\mathbf{W} = (\mathbf{u}, -1)$ com
 $\mathbf{W}_i = (U_i, -1)$
- $\pi(d_j|U_i, Z_{ij}^*, V_{ij}, \beta_j, D_{obs}) \propto \phi\left(\frac{m_{ij} - Z_{ij}^*}{V_{ij}}, \frac{1}{V_{ij}^2}\right)$.

Prova:

Da verossimilhança dada em (4.13) e das distribuições a priori dadas na Seção 4.4.1, temos que a distribuição conjunta a posteriori é dada por:

$$f(\mathbf{u}, \beta, \mathbf{d}, \mathbf{Z}^*, \mathbf{X}) = \prod_{i=1}^n \prod_{j=1}^I \phi(Z_{ij}^*; -d_i V_{ij} + m_{ij}, 1) I(Z_{ij}^*, y_{ij}) \phi(V_{ij}; 0, 1) I(V_{ij} > 0)$$

$$\prod_{i=1}^n \phi_{NA}(U_i, \kappa) \prod_{j=1}^I \phi_2(\beta_j, \mu_\beta, \Sigma_\beta)$$

A prova segue substituindo a densidade $\phi_{SN}(\cdot)$ pela expressão (2.2) e identificando na distribuição conjunta os termos que contêm os parâmetros desejados na condicional, e agrupando as densidades correspondentes $\phi(\cdot)$ por um lado e considerando a distribuição acumulada $\Phi(\cdot)$

por outro lado. As condicionais de Z_{ij}^* , U_i e β_j são as mesmas que as dadas por Johnson e Albert (1999) para o caso do modelo PN em que não é considerado assimetria na variável latente e na função de ligação. \square

4.4.4 Implementação do MCMC

Nós desenvolvemos um procedimento computacional para a família TRI-NA baseados na segunda versão da função de verossimilhança aumentada dada em (4.1.4). Semelhante ao caso da regressão binária com ligação probito assimétrica desenvolvida no Capítulo 3, duas parametrizações serão possíveis: d e λ . A especificação da verossimilhança completa para a *parametrização* d é dada por:

$$\begin{aligned} Z_{ij}^* | x_{ij}, y_{ij}, a_j, b_j, d_j &\sim N(m_{ij} - d_j x_{ij}, 1 - d_j^2) I(z_{ij}^*, y_{ij}); \\ V_{ij} &\sim HN(0, 1); \\ u_i &\sim NA(\theta_u), \theta_u = (0, 1, \kappa); \\ a_j &\sim HN(\mu_a, \sigma_a); \\ b_j &\sim N(0, \sigma_b^2); \\ &e \\ d_j &\sim U[-1, 1]. \end{aligned}$$

A especificação de verossimilhança completa para a *parametrização* λ é dada por:

$$\begin{aligned} Z_{ij}^* | y_{ij}, a_j, b_j, \lambda_j &\sim N\left(m_{ij} - \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}} V_{ij}, \frac{1}{1 + \lambda_j^2}\right) I(Z_{ij}^*, y_{ij}); \\ V_{ij} &\sim HN(0, 1); \\ u_i &\sim NA(\theta_u), \theta_u = (0, 1, \kappa); \\ a_j &\sim HN(\mu_a, \sigma_b); \\ b_j &\sim N(0, \sigma_b^2); \\ &e \\ \lambda &\sim T(\mu, \sigma^2, v). \end{aligned}$$

Esta estrutura hierárquica pode ser facilmente introduzida no programa WinBUGS. Note que quando $d_j = 0$, a estrutura hierárquica da verossimilhança aumentada correspondente ao modelo probito-normal segue eliminando a segunda e quinta linha na hierarquia acima.

4.5 Probabilidade não condicional de resposta correta sob a família TRI-NA

A seguinte propriedade da família TRI-NA é uma generalização da propriedade do modelo PN simétrico.

Proposição 10. Considere a família TRI-NA com parâmetros de item β_j e d_j e denotemos a probabilidade condicional de resposta correta com uma variável latente u , como $p_{ij} = P(Y_{ij} = 1|u_i, \beta_j) = p_j(u)$, então, a probabilidade de resposta correta não condicional para o item j é dada por

$$p_j = \int_{-\infty}^{\infty} P_j(u)\phi_{SN}(u; \kappa)du = 4\Phi_3 \left[\left(\begin{array}{c} -b_j \\ 0 \\ 0 \end{array} \right); \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right), \left(\begin{array}{ccc} 1 + a_j^2 & -d_j & a_j\kappa \\ -d_j & 1 & 0 \\ a_j\kappa & 0 & 1 + \kappa^2 \end{array} \right) \right]$$

Prova: A prova usa propriedades da distribuição normal dada por Liseo e Loperfido (2003) que é útil para o cálculo com distribuições normais assimétricas (ver apêndice 1). Argumentos similares como os dados em Albert (1992) para o caso do usual modelo PN podem ser usados. \square

Dois casos especiais são de interesse. Se nós consideramos assimetria unicamente para a CCI ($\kappa = 0$), o modelo obtido é o PAN, e então

$$p_j = \int_{-\infty}^{\infty} P_j(u)\phi(u)du = 2\Phi_2 \left[\left(\begin{array}{c} -b_j \\ 0 \end{array} \right); \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 + a_j^2 & -d_j \\ -d_j & 1 \end{array} \right) \right] = \Phi_{NA} \left[\frac{-b_j}{\sqrt{1 + a_j^2}}; 0, 1, -\frac{-d_j}{\sqrt{1 + a_j^2 + d_j^2}} \right]$$

Por outro lado, considerando assimetria só na variável latente (i.e $d_j = 0$), o modelo obtido é o modelo PNA com

$$p_j = \int_{-\infty}^{\infty} P_j(u)\phi_{SN}(u; \kappa)du = 2\Phi_2 \left[\left(\begin{array}{c} -b_j \\ 0 \end{array} \right); \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 + a_j^2 & a_j\kappa \\ a_j\kappa & 1 + \kappa^2 \end{array} \right) \right]$$

As duas expressões acima apresentam generalizações com respeito ao usual modelo PN porque se não existe assimetria na CCI e na variável latente nós obtemos a probabilidade não condicional de obter uma resposta correta para o item j é dada por $p_j = \Phi\left(\frac{-b_j}{\sqrt{1+a_j^2}}\right)$ como é apresentado em Albert (1992). Esta é a probabilidade de um indivíduo aleatoriamente escolhido obtenha uma resposta correta para o item j .

Capítulo 5

Aplicações da família de modelos TRI-NA

5.1 Introdução

Neste capítulo, implementamos a inferência Bayesiana para a família de modelos TRI-NA desenvolvida no capítulo 4 em dois conjuntos de dados. Um conjunto apresentado em Tanner (1996, p. 190) é usado para um estudo da eficiência na estimação dos parâmetros do modelo, usando a noção de tamanho efetivo (ESS) da amostra como definido em Kass *et al.* (1998) e o tamanho de amostra por segundo (ESS/s) como considerado em Sahu (2002). A metodologia proposta é ilustrada através de um conjunto de dados de um teste de Matemática aplicado a estudantes de escolas peruanas de ensino básico. Estes dados já foram utilizados no Capítulo 2 ao se considerar a o modelo PN. Para este último conjunto de dados também é analisado brevemente a sensibilidade do modelo PAN à escolha de diferentes prioris para os parâmetros de discriminação e dificuldade.

Nós também mostramos a interpretação do novo parâmetro proposto e mostramos as vantagens do modelo PAN sobre o modelo PN ao se considerar a probabilidade de resposta não condicional para o itens .

5.1.1 A implementação da estimação MCMC da TRI-NA

A seguir são discutidas a especificação de prioris, os valores iniciais para definir os estados iniciais da cadeia de Markov, e diagnósticos de convergência para cadeia de Markov, na implementação do algoritmo MCMC. O procedimento MCMC esta baseado nos resultados da Proposição 4 e é implementado no programa WinBUGS.

A especificação de prioris

Como tem sido dito, para a família de modelos TRI-NA é preciso especificar prioris próprias para garantir que a posterior seja própria. Também Albert e Ghosh (2000) observam que a escolha de prioris próprias para os traços latentes resolvem o problema de identificação do modelo. Além disso, distribuições a priori informativas podem ser especificadas e podem ser usadas para refletir o fato de se acreditar a priori que os valores dos parâmetros dos itens não são extremos (isto é, não estão na fronteira do espaço paramétrico). Nas situações comuns quando

pouca informação a priori é disponível a respeito dos parâmetros de dificuldade, nós podemos escolher uma variância grande, isto é σ_b^2 grande. Esta escolha pode ter um efeito modesto na distribuição posterior para dados não extremos, e pode resultar em uma distribuição posterior própria quando dados extremos (em que estudantes apresentam todas as respostas ao itens corretas ou incorretas) são observados (Albert e Ghosh, 2000). Além disso, Sahu (2002) estabelece que a especificação de grandes valores da variância de b levam a estimativas instáveis. No Capítulo 2 nós comparamos o uso e a performance de seis diferentes prioris dadas na literatura para os parâmetros de discriminação e dificuldade no modelo PN, através de uma análise de sensibilidade. Isto incluiu a especificação de distribuições a priori vagas para os parâmetros de dificuldade e prioris precisas para o parâmetro de discriminação. As prioris especificadas foram sugeridas em pesquisas prévias por outros autores. As diferentes prioris consideradas levaram a similares estimativas do *DIC* (descrito em Spiegelhalter *et al.*, 2002) permitindo concluir que para o conjunto de dados considerados a análise bayesiana não é sensível escolha da priori. Embora, a nosso entender, as prioris consideradas em Sahu (2002), a qual especifica que $a_j \sim N(1; 0.5)I(a_j > 0;)$ e $b_j \sim N(0; 2), j = 1, \dots, k$, resultaram mais adequadas porque se obtêm estimativas de grande precisão para os parâmetros de discriminação e dificuldade. Para os outros modelos com ligação probito assimétrica, i.e., modelos PAN e PANA, as prioris foram especificadas para d_j e não (diretamente) para $\lambda_j, j = 1, \dots, k$. Nós escolhemos trabalhar com o parâmetro de assimetria d por oferecer a mesma interpretação que o parâmetro λ numa escala de valores em $[-1, 1]$. A priori especificada para d_j é $U(-1, 1)$ ou equivalentemente $\lambda_j \sim T(0, 0.5, 2)$. Para os modelos com assimetria na variável latente, i.e. modelos PANA e PNA, nós consideramos $U_i \sim NA(\kappa), i = 1, \dots, n$. Nós consideramos para o modelo PANA, uma especificação baseada nos valores sugeridos pelos escores no teste os quais para o conjunto de dados sob análise (veja-se seção 5.3) levam a $\mu_\kappa = 0.804$ e $\sigma_\kappa^2 = 6.329$. Esta especificação esta baseada em que a distribuição dos escores é uma aproximação grosseira da distribuição das variáveis latentes.

Valores iniciais

Para a TRI-NA nós consideramos, como em Spiegelhalter (1996), valores iniciais de 1 e 0 para os parâmetros de item a_j e $b_j, j = 1, \dots, k$, respectivamente. Para o modelo PANA nós consideramos o valor de 0 para o parâmetro κ . Para os modelos PAN e PANA valores iniciais para o parâmetro de assimetria d_j podem ser aleatoriamente gerados, entretanto, preferimos fixa-los. Nós propomos valores iniciais $d_j = 0$ or $\lambda_j = 0$ para o parâmetro de assimetria porque isto corresponde ao valor médio o valor esperado da distribuição uniforme em $[-1, 1]$. Valores iniciais para U_i as variáveis latentes e auxiliares tais como Z_{ij}^* e V_{ij} são aleatoriamente geradas considerando as correspondentes distribuições.

Convergência da cadeia de Markov

O modelo PANA é o mais geral e para o conjunto de dados do teste de Matemática ele envolve 42 parâmetros de item e 131 traços latentes (para 131 indivíduos na amostra). Embora nosso interesse seja restrito por simplicidade a média e desvio padrão dos traços latentes da amostra antes do que as estimativas dos traços latentes ainda o número de parâmetros é grande ($42 + 2$). Por sua vez, ao se considera a metodologia DAGS para MCMC como já foi observado no capítulo 2 e 3, um grande número de cadeias deve ser gerado para os diferentes parâmetros, a que torna o procedimento MCMC demorado (com o algoritmo considerado, o modelo PN toma aproximadamente 1 minuto para executar 1000 iterações num processador Pentium IV com 256 MB RAM). De maneira informal, a convergência no modelo PAN leva o dobro de tempo que leva o modelo PN; já no modelo PANA leva aproximadamente 1.5 vezes o tempo que leva no modelo PN, e a geração dos valores foi um pouco mais demorada sob o modelo PANA mas menor do que leva no modelo PN.

O tempo necessário para a execução das cadeias está também relacionada com a presença de estruturas das variáveis latentes (Chen *et al.* 2000), e com o tamanho da amostra (Sahu, 2002), o qual pode afetar significativamente o tempo de execução dos modelos TRI-NA como já foi observado no capítulo 2 e 3. Quando é usado o MCMC, os valores amostrais para as iterações iniciais das cadeias são descartados devido a sua dependência dos valores iniciais. Também, com modelos TRI-NA, presença de autocorrelações entre os valores da cadeia é esperado devido a introdução de variáveis latentes auxiliares (Chen *et al.* 2000). Nossa experiência recomenda, valores de salto acima de 100 e o descarte de 2000 a 4000 iterações iniciais. Vários critérios calculados usando o programa CODA e BOA no *R* incluindo o proposto por Geweke (1992) foram usadas para análise de convergência. Uma alternativa para ser considerada é gerar um grande número de iterações e usar valores grandes de salto.

5.1.2 Critérios de Comparação de Modelos

Para comparar os diferentes modelos da TRI-NA, nós estimamos o valor *esperado do desvio bayesiano posterior* (D_{bar}), o *desvio bayesiano das médias a posteriori* (D_{hat}), o *número efetivo de parâmetros* ρ_D e o *critério de informação de desvio* (DIC), como apresentado em Spiegelhalter *et al.* (2002). Como também foi observado no Capítulo 2, na presença de variáveis latentes auxiliares, DIC s marginais para as variáveis observadas devem ser especificados devido ao fato de que o foco de análise é a verossimilhança $p(\mathbf{y}|\mathbf{u}, \mathbf{a}, \mathbf{b}, \mathbf{d})$ antes do que a verossimilhança aumentada $p(\mathbf{y}, \mathbf{Z}^*, \mathbf{V}|\mathbf{u}, \mathbf{a}, \mathbf{b}, \mathbf{d})$ na metodologia DAGS. Além disso, como no Capítulo 3, a comparação dos modelos da TRI-NA é feita usando o *Critério de Informação de Akaike* esperado ($EAIC$) e o *Critério de informação (Schwarz) bayesiana* esperado ($EBIC$) como sugerido em Carlin e Louis (2000) e Brooks (2002), e a *soma de quadrados de residuais latentes a posteriori* ($SQRL$). Para maiores detalhes veja o Apêndice B.

5.2 Um estudo da eficiência na estimação dos parâmetros da família TRI-NA

Nesta seção, nós consideramos, um estudo da eficiência na estimação dos parâmetros da família TRI-NA usando um conjunto de dados previamente analisados na literatura. Os propósitos desta pesquisa são: a) avaliar a conduta da autocorrelação dos parâmetros do modelo quando o algoritmo dos dados aumentados é considerado, e b) avaliar o uso das *parametrizações* d e λ introduzidas na seção 4.4.4. Consideramos o conjunto de dados apresentado em Tanner (1996, pág 190), o qual é constituído de $k = 6$ ítems e $n = 39$ indivíduos.

Como já foi dito, é esperado a presença de autocorrelações entre valores das cadeias para a TRI-NA devido ao uso de técnicas de dados aumentados. Considerando este fato, é conveniente executar uma análise das autocorrelações para os parâmetros. Restringimos este estudo ao modelo PAN a ao modelo PN. Com o propósito de fazer comparações entre a eficiência na estimação dos parâmetros propomos usar a noção de *tamanho de amostra efetiva* (effective sample size : ESS) como definido em Kass *et al.* (1998) e o *tamanho de amostra por segundo* (ESS/s) como considerado em Sahu (2002), isto é, ESS dividido pelo tempo de execução. ESS é definido para cada parâmetro como o número de amostras MCMC, dividido pelo tempo de autocorrelação, $\gamma = 1 + 2 \sum_{s=1}^{\infty} \rho_s$, em que ρ_s é a autocorrelação até o lag s . Estimação de γ usando autocorrelações amostrais é problemático porque amostras MCMC são usadas na estimação ρ_s quando s cresce. Nós usamos $s = 49$ como mostrado no WinBUGS pois para valores $s \geq 50$ a estimativa é desprezível. Sahu (2002) usaram ESS/s para comparar dois diferentes aproximações de estimação para o mesmo modelo e sinalizam que a taxa $EF_{12} = \frac{ESS/s_1}{ESS/s_2}$, pode ser considerada como uma medida de eficiência entre as aproximações de estimação, de maneira que grandes valores de EF_{12} indicam que a estimação é mais eficiente sob a aproximação 1 antes do que sob a aproximação 2. Nós propomos usar este índice para comparar a eficiência na estimação de parâmetros diferentes sob o mesmo modelo e para o mesmo parâmetro estimado sob modelos diferentes. Nós usamos este conceito para estudar a eficiência na estimação de um mesmo parâmetro entre o modelo probito assimétrico-normal (PAN) e o modelo probito-normal (PN). Este conceito também é usado na comparação da eficiência na estimação dos parâmetros do modelo PAN quando considerada a *parametrização* d ou a *parametrização* λ . Todos estes casos são considerados neste análise como modelos diferentes a serem comparados usando EF_{12} .

Na Tabela 5.1 apresentamos a média e desvio padrão das estimativas de ESS e ESS/s por cada tipo de parâmetro (obtidas sob todos os ítems ou sob todos os indivíduos segundo seja o caso) no modelo PN e no modelo PAN nas *parametrizações* d ou λ . Usando o índice de eficiência definido acima, a Figura 5.1 mostra intervalos de confiança ao 95% para a Eficiência (EF_{12}) de cada parâmetro para cada dois modelos considerados. Considerando os resultados da Tabela 5.1 e a figura 5.1 notamos que:

Tabela 5.1 Performance na estimação de parâmetros para os dados do exemplo de Tanner (Dp: desvio padrão)

Modelo	Parâmetro	ESS		ESS/s	
		Médio	Dp	Média	Dp.
PAN (parametrização d)	a	260,7	125,9	3,7	1,8
	b	157,6	20,9	2,2	0,3
	d	137,3	16,7	1,9	0,2
	λ	121,4	13,7	1,7	0,2
	u	582,9	165,5	8,2	2,3
	Total		425,3	245,6	6,0
PAN (parametrização λ)	a	182,8	93,8	2,3	1,1
	b	100,2	9,7	1,2	0,1
	d	76,4	7,6	0,9	0,1
	λ	71,1	8,4	0,9	0,1
	u	509,4	199,9	6,3	2,5
	Total		356,4	254,4	4,4
PN	a	307,9	127,5	9,1	3,7
	b	905,2	338,8	26,6	10,0
	u	783,5	236,5	23,0	7,0
	Total		741,9	287,7	21,8

- Para qualquer parametrização (d ou λ) no modelo PAN existe um ordenamento na eficiência na estimação dos parâmetros do modelo. A estimação é mais eficiente para o parâmetro u , seguida pelos parâmetros a , os quais são seguidos pelos parâmetros b , e, finalmente, d seguido dos parâmetros λ .
- No modelo PAN, a parametrização d é muito mais eficiente na estimação dos parâmetros d e λ do que a parametrização λ . Além disso, na mesma parametrização, a estimação d é mais eficiente que a estimação λ .
- Quando um modelo PN é mais apropriado para um conjunto de dados (como é o caso do conjunto de dados analisados), nota-se que existe uma redução da eficiência na estimação

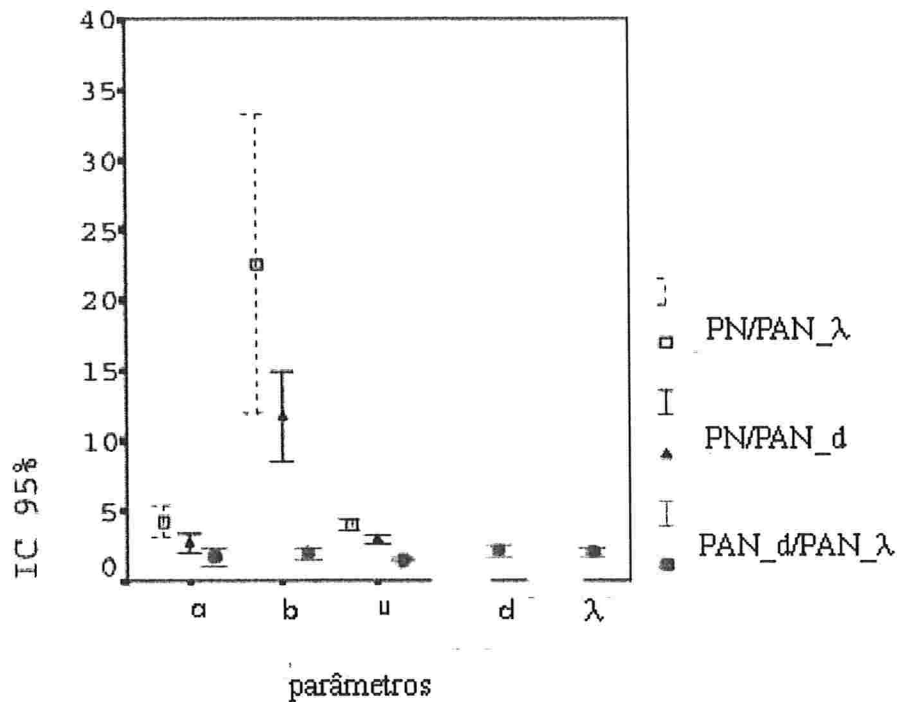


Figura 5.1 Comparação da performance das estimativas dos parâmetros para o exemplo de Tanner usando intervalos de credibilidade de 95 % da eficiência entre os modelos considerados

do parâmetro a , b , u quando o modelo PAN é considerado, especialmente quando usamos a parametrização λ .

Outros conjunto de dados (os conhecido LSAT e o exemplo de Matemática na seguinte seção) apresentam uma conduta similar à descrita acima.

Considerando a presença das autocorrelações no modelo PAN, valores de salto acima de 100 são recomendados. Conseqüentemente uma grande número de iterações são necessárias para a inferência baseada na densidade conjunta a posteriori. Este tipo de conduta foi também observada para outros modelos. Jackman (2004) considerando o modelo PN, executou meio milhão de iterações e só uma de cada mil destas iterações foram consideradas para produzir uma seqüência independente de valores amostrais da densidade conjunta a posteriori.

Chen *et al.* (2000) alertam que quando o tamanho da amostra n é grande ($n \geq 50$), é esperada

convergência demorada do algoritmo de Albert-Chib. Na TRI-NA é encontrada convergência demorada da cadeia para o parâmetro de assimetria especialmente na parametrização λ . Alguns algoritmos para melhorar a convergência do amostrador de Gibbs na segunda abordagem do algoritmo de dados aumentados são sugeridas em Chen *et al.* (2001) e devem ser exploradas em trabalhos subseqüentes.

5.3 Uma aplicação para um Teste de Matemática

Na pesquisa em avaliação educacional é típico detectar diferenças na proficiência escolar devido ao nível sócio econômico. Por exemplo, numa pesquisa no Perú, Bazán *et al.* (2001) reportam diferenças observadas num teste de matemática para alunos de sexta série em favor dos alunos com alto nível sócio econômico. Isto é, a maioria de alunos com alto nível sócio econômico tendem a obter altos escores no teste e porém os escores do teste mostram uma distribuição assimétrica.

Nós ilustramos a abordagem bayesiana da família TRI-NA usando os dados correspondentes a um teste de matemática aplicado em escolas peruanas com o propósito de estimar os parâmetros dos ítems, incluindo o novo parâmetro de assimetria. Pretendemos mostrar as vantagens do modelo PAN com respeito ao modelo PN para capturar mais informação do item na presença de escores assimétricos.

5.3.1 Os dados

O conjunto de dados corresponde à aplicação de 14 ítems de um teste de matemática disponíveis para “download” em <http://www.minedu.gob.pe/umc/> aplicado para 131 alunos de alto nível sócio econômico da sexta série.

Nesta aplicação, 14 ítems do teste de Matemática do ensino básico (UMC, 2001) foram aplicados a 131 alunos de alto nível sócio econômico. Vetores de resposta aparecem no Apêndice D e estatísticas resumo são apresentadas na Tabela 2.1 do Capítulo 2. Este mesmo conjunto de dados já foi analisado no Capítulo 2 sob modelo PN. Como fora indicado, a distribuição dos escores apresenta uma assimetria de -0.804 e uma proporção média de resposta correta para os ítems de 0.774, indicando que o teste pode ser considerado fácil nesta população. Isto mostra que os escores dos alunos de sexta série apresentam assimetria negativa; de aqui gráficos do escore do item vs escore do teste (regressão ítem teste) para diferentes ítems não apresentam forma simétrica. Embora uma regressão ítem-teste não seja uma aproximação fechada como a curva característica do item (Lord e Novick, 1968, p. 363), isto pode dar ideia da verdadeira forma da CCI. Isto justifica explorar os modelos da família TRI-NA para este conjunto de dados.

5.3.2 Uma comparação de modelos PAN e PN

Com o propósito de nos familiarizarmos com o modelo PAN nós usamos duas diferentes combinações das distribuições a priori para os parâmetros tradicionais dos ítems (dificuldade e

discriminação) já usadas no modelo PN. Elas são:

Priori A: $a_j \sim N(0, 1)I(a_j > 0)$, $b_j \sim N(0, 10000)$, $d_j \sim U(-1, 1)$ e

Priori B: $a_j \sim N(1, 0.5)I(a_j > 0)$, $b_j \sim N(0, 2)$, $d_j \sim U(-1, 1)$.

A priori B é uma priori mais precisa e a priori A tem uma priori difusa para o parâmetro b . A para d e u já foi indicada na Seção 4.1.

O modelo PAN envolve 42 parâmetros de item: $(a_j, b_j, \lambda_j(d_j))$, $j = 1, \dots, 14$ e 131 parâmetros das variáveis latentes $(U_i, i = 1, \dots, 131)$. Embora este procedimento permita obter as distribuições a posteriori dos traços latentes individuais para os 131 alunos na amostra nós reportamos apenas a média e desvio padrão destes traços latentes. Como indicado na Seção 5.1.1, um grande número de iterações deve ser gerado para o modelo devido a presença de autocorrelação, o que torna o procedimento MCMC demorado. Para propósitos de inferência a respeito dos quatro possíveis modelos propostos, nós geramos 304000 iterações e descartamos os primeiros 4000 valores iniciais. Usando um salto de 150, o tamanho efetivo de amostras consideradas foi de 2000. A parametrização d foi usada. Estimativas dos parâmetros do modelos foram calculados usando essas iterações.

Para avaliar a convergência vários critérios foram considerados usando o programa CODA, como por exemplo, o proposto por Geweke (1992).

Para a comparação dos modelos PAN e PN utilizou-se apenas o DIC . Considerando-se os valores mostrados na Tabela 2, concluímos que o modelo PAN, melhora o correspondente modelo PN simétrico para qualquer priori considerada. Spiegelhalter *et al.* (2002) diz que ρ_D pode ser negativa e isto pode indicar conflito entre as prioris especificadas e os dados ou que a média posterior é um pobre estimativa. Nos achamos que a mediana posterior pode ser melhor estimativa especialmente no parâmetro d desde que é observado uma distribuição assimétrica neste parâmetro. Também a especificação de prioris informativas usando dados históricos, como os propostos por Chen, *et al.* (2001) podem ser explorados em subseqüentes trabalhos.

Tabela 5.2 Comparação dos modelos TRI PAN e PN usando DIC

modelos	parâmetros	$Dbar$	$Dhat$	ρ_D	DIC
PN	159	1446,17	1358,13	88,04	1534,21
PAN (prioris A)	173	1315,88	1370,39	-54,511	1261,37
PAN (prioris B)	173	1330,97	1367,5	-36,528	1294,44

As análises do Capítulo 2, usando as prioris A e B, indicaram que o modelo PN é insensível a especificação destas prioris. Entretanto, este resultado não é observado no modelo PAN, o qual é sensível à especificação de prioris difusas (menos informativas) como veremos a seguir.

A Tabela 5.3 mostra correlações estimadas entre as médias a posteriori dos correspondentes parâmetros para as prioris consideradas. Note que todos os parâmetros de item são correlacionados sob a priori B. Note que, para as prioris A e B, os parâmetros a e b apresentam uma forte correlação entre elas. Entretanto, é importante observar que d (ou λ) sob as duas prioris apresentam uma correlação muito baixa. Isto é, não são correlacionados. As estimativas de d ou λ são diferentes quando as prioris A e B são consideradas. Quando a priori A é usada as médias a posteriori de d (ou λ) estão todas em torno de zero não permitindo identificar assimetria como acontece com a priori B. Nós escolhemos apresentar os resultados de inferência considerando a priori B.

Tabela 5.3 *Correlação das médias a posteriori para os parâmetros dos item sob duas tipos de prioris para os parâmetros de discriminação e dificuldade*

		priori A				priori B			
		a	b	d	λ	a	b	δ	λ
priori A	a	1							
	b	-0,71	1						
	d	-0,31	-0,32	1					
	λ	-0,23	-0,39	0,92	1				
priori B	a	0,99	-0,70	-0,31	-0,22	1			
	b	-0,64	0,99	-0,38	-0,45	-0,65	1		
	d	-0,90	0,88	0,09	0,02	-0,87	0,83	1	
	λ	-0,91	0,74	0,27	0,19	-0,87	0,67	0,97	1

Estimativas dos parâmetros de ítem discriminação e dificuldade para os modelos PAN e PN usando as priori B são apresentados na Figura 5.2. Os dois tipos de parâmetros são igualmente interpretáveis sob ambos modelos. O ítem 11 é mais discriminativo e o ítem 9 é o menos discriminativo. Também, o ítem 11 é mais fácil e o ítem 12 é o mais difícil. Assim, o modelo PAN é um modelo que oferece as mesmas conclusões para os parâmetros de dificuldade e discriminação do que o modelo PN. Na Figura 5.3 nós apresentamos as diferenças entre os modelos PN e PAN (soma de valores absolutos das diferenças nas médias a posteriori dos parâmetros a_j e b_j) para o novo parâmetro de ítem d_j .

Como nós esperamos, as diferenças nos modelos PN e PAN são mínimas (os parâmetros de dificuldade e de discriminação nos modelos PN e PAN são aproximadamente iguais) quando o parâmetro de assimetria se aproxima de zero. Isto não acontece no caso dos ítems 11 e 4, os quais

apresentam estimativas dos parâmetros de penalidade (d ou λ) altos e negativos (assimetria negativa na CCI). Nestes casos, os parâmetros de dificuldade e discriminação diferem nos dois modelos. No caso especial dos itens 11 e 4, a diferença entre modelos como consequência do parâmetro de assimetria afeta o parâmetro de dificuldade. Os outros itens apresentam estimativas dos parâmetros de penalidade perto de zero indicando que o modelo PN é mais adequado para explicar a conduta destes itens.

5.3.3 O uso da família TRI-NA

Para ilustrar o uso da família TRI-NA e sua utilidade, os modelos PN e PAN foram ajustados novamente para ser comparados com o ajuste do modelo PANA. Neste caso nós consideramos vários critérios de comparação apresentados nos capítulos anteriores e descritos no Apêndice B.

Aqui nós consideramos um total de 204000 iterações para todos os modelos. Iniciando com um burn-in de 4000 iterações e então usando salto igual a 100, é obtido um tamanho de amostra de 2000.

Tabela 5.4 Comparando os modelos PN, PAN e PANA usando diferentes criterios

Criterios	Modelos TRI-NA		
	Modelo PN	Modelo PAN	Modelo PANA
Número de parâmetros	159	173	174
$Dbar$	1446,17	1317,28	1321,83
$Dhat$	1358,13	1365,11	1363,35
ρ_D	88,04	-47,83	-41,52
DIC	1534,21	1269,44	1280,32
$EAIC$	1764,17	1663,28	1669,83
$EBIC$	2640,94	2617,25	2629,31
SQRL	1853	1362	1351

Considerando os valores DIC , $EAIC$, $EBIC$ na Tabela 5.4, nós vemos que fica confirmado que o modelo PAN melhora o correspondente modelo PN simétrico e o modelo PANA. Este último modelo também apresenta melhor ajuste do que o modelo PN. Este resultado é também observado considerando a soma de quadrados dos residuais latentes (SQRL) (ver Figura 5.4), no qual o modelo PANA pode ser alternativo ao modelo PAN para o conjunto de dados. Resumindo, o modelo PAN apresenta o melhor ajuste para o conjunto de dados.

5.3.4 A interpretação dos parâmetros

Assumindo que o modelo PAN é o melhor modelo para os dados propostos nós realizamos análises adicionais a respeito da interpretação dos parâmetros deste modelo em relação com o modelo PN.

As Tabelas 5.5 e 5.6 ilustram a conduta das estimativas dos parâmetros dos ítems para os modelos PN e PAN. Com já foi dito, estimativas dos parâmetros de discriminação e dificuldade para ambos modelos apresentam conduta similar como pode ser analisado na tabela 5.5. Os dois tipos de parâmetros são igualmente interpretáveis sob os dois modelos como fora discutido na Secao 4.2.3. O modelo PAN é um modelo que oferece as mesmas conclusões que o modelo PN a respeito dos parâmetros de dificuldade e discriminação porque estes são igualmente interpretáveis.

A Tabela 5.5 mostra os valores estimados do parâmetro de assimetria ou penalidade d_j . A Figura 5.5 apresenta os histogramas das distribuições a posteriori do parâmetro d , o qual indica a presença de densidades simétricas e assimétricas. É importante dizer que considerando a noção de tamanho de amostra efetivo (ESS) usada em Sahu (2002) nós encontramos que o parâmetro a tem melhor convergência. A convergência para os parâmetros b é mais demorada, uma indicação que uma cadeia grande é necessária. Isto também acontece com os parâmetros d , mas em menor grau. Existe uma indicação que se obtém melhor convergência com a parametrização d do que com a parametrização λ .

Na Tabela 5.6, também é apresentada as diferenças entre a média a posteriori para os parâmetros a_j e b_j nos modelos PN e PAN. Como esperávamos, os parâmetros de dificuldade e discriminação são aproximadamente iguais quando o parâmetro de assimetria se aproxima a zero. Como estimador do parâmetro de penalidade d_j resulta mais conveniente considerar a mediana a posteriori pela assimetria da distribuição a posteriori. Ítems 11, 4, e 7 são os únicos que apresentam estimativas do parâmetro de penalidade d_j grandes e negativos. Portanto os parâmetros de discriminação e dificuldade diferem nos dois modelos. A probabilidade a posteriori de valores negativos para o parâmetro de assimetria foi calculada para os ítems 11, 4, e 7, resultando nos seguintes valores: 0,732, 0,614 e 0,632, respectivamente, o qual indica que esses parâmetros não são iguais a zero, indicando que o uso de CCI assimétricas para estes ítems é adequado. No caso especial dos ítems 11, 4, e 7, as diferenças entre os modelos como consequência do parâmetro de assimetria afetam o parâmetro de dificuldade. Também observamos que os ítems 3 e 6 apresentam estimativas positivas dos parâmetros de penalidade indicando que eles apresentam CCI com assimetria positiva mas seus valores não são significativamente diferentes de zero, e também que outros ítems podem ser modelados corretamente considerando o modelo PN com CCI simétricos.

O ítem 11 diz: “Luiza, Dora e Maria compram certa quantidade de tela. Luiza comprou a metade de um metro, Dora comprou 75 centímetros e Maria comprou 50 centímetros. Qual delas comprou a mesma quantidade de tela?”. O ítem 4 diz: “Pepe divide um número por 17, obtendo um quociente de 9 e um resto de 2. Qual é o número dividido?”. Os valores altos e negativos para

Tabela 5.5 Média posterior, desvio padrão (dp) e intervalos HPD de 90 % para os parâmetros comuns sob os modelos PN e PAN

	Modelo PN					Modelo PAN				
	média	dp	mediana	90 % HPD		média	dp	mediana	90 % HPD	
				Inferior	Superior				Inferior	Superior
a1	0,535	0,228	0,520	0,149	0,880	0,491	0,222	0,469	0,147	0,860
a2	0,300	0,181	0,280	0,007	0,557	0,269	0,172	0,245	0,001	0,499
a3	0,552	0,228	0,531	0,197	0,912	0,496	0,209	0,480	0,141	0,809
a4	0,915	0,348	0,889	0,332	1,462	0,856	0,351	0,824	0,252	1,375
a5	0,512	0,249	0,493	0,117	0,912	0,465	0,231	0,440	0,093	0,824
a6	0,314	0,166	0,299	0,026	0,551	0,284	0,154	0,267	0,040	0,524
a7	0,838	0,341	0,801	0,307	1,375	0,801	0,341	0,755	0,251	1,339
a8	0,959	0,350	0,928	0,350	1,470	0,885	0,331	0,858	0,341	1,401
a9	0,199	0,140	0,171	0,001	0,393	0,180	0,130	0,158	0,000	0,355
a10	0,496	0,237	0,479	0,112	0,879	0,451	0,233	0,430	0,053	0,782
a11	1,323	0,394	1,296	0,662	1,915	1,266	0,435	1,232	0,559	1,928
a12	0,387	0,196	0,369	0,043	0,668	0,357	0,179	0,340	0,047	0,616
a13	0,457	0,225	0,432	0,096	0,801	0,405	0,212	0,379	0,048	0,726
a14	0,445	0,276	0,411	0,001	0,825	0,404	0,257	0,364	0,002	0,753
b1	-0,900	0,153	-0,895	-1,124	-0,630	-0,780	0,421	-0,817	-1,449	-0,123
b2	-1,083	0,144	-1,080	-1,308	-0,840	-0,976	0,426	-1,043	-1,613	-0,285
b3	-0,026	0,127	-0,028	-0,217	0,195	-0,049	0,394	-0,055	-0,691	0,567
b4	-1,902	0,338	-1,861	-2,409	-1,374	-1,662	0,581	-1,701	-2,591	-0,690
b5	-1,232	0,178	-1,218	-1,524	-0,952	-1,057	0,458	-1,124	-1,798	-0,337
b6	0,373	0,121	0,371	0,181	0,570	0,328	0,392	0,347	-0,304	0,933
b7	-1,779	0,317	-1,746	-2,267	-1,278	-1,519	0,561	-1,578	-2,397	-0,574
b8	-1,517	0,294	-1,476	-1,953	-1,045	-1,332	0,512	-1,396	-2,060	-0,417
b9	-0,794	0,131	-0,790	-1,010	-0,592	-0,731	0,406	-0,777	-1,330	-0,079
b10	-1,179	0,176	-1,166	-1,467	-0,898	-1,051	0,441	-1,112	-1,709	-0,321
b11	-2,290	0,440	-2,236	-3,025	-1,663	-1,829	0,719	-1,862	-2,945	-0,586
b12	0,424	0,123	0,418	0,226	0,627	0,388	0,386	0,405	-0,199	1,021
b13	-0,995	0,156	-0,988	-1,230	-0,740	-0,884	0,438	-0,934	-1,526	-0,174
b14	-1,726	0,239	-1,705	-2,111	-1,388	-1,522	0,502	-1,612	-2,306	-0,736
média u	0,045	0,087	0,045	-0,105	0,1819	0,036	0,087	0,039	-0,109	0,178
dp u	0,933	0,062	0,933	0,872	1,039	0,931	0,060	0,930	0,828	1,027

as estimativas dos parâmetros de penalização d indicam que estes itens penalizam aos alunos com maior conhecimento e bonificam aos estudantes com menor conhecimento. Alunos com alto conhecimento têm poucas diferenças nas suas probabilidades de sucesso para o item, mais para estudantes com menor conhecimento, possivelmente um esclarecimento do fato que a metade do metro equivale a 50 centímetros no enunciado do item pode produzir uma mudança significativa na probabilidade de sucesso.

Tabela 5.6 Média, mediana e desvio padrão (dp) a posteriori e intervalos HPD de 90 % para os parâmetros d sob o modelo PAN

	Estatísticas a posteriori					Diferenças entre PN e PAN	
	média	dp	mediana	90 % HPD		Diferença média	
				Inferior	Superior	En a	En b
d1	-0,058	0,486	-0,084	-0,772	0,823	0,044	-0,120
d2	-0,021	0,499	-0,053	-0,787	0,843	0,031	-0,108
d3	0,018	0,489	0,010	-0,802	0,781	0,056	0,023
d4	-0,110	0,525	-0,175	-0,892	0,788	0,059	-0,240
d5	-0,089	0,488	-0,136	-0,896	0,665	0,046	-0,175
d6	0,006	0,488	0,027	-0,849	0,730	0,030	0,044
d7	-0,145	0,516	-0,219	-0,975	0,640	0,037	-0,260
d8	-0,077	0,506	-0,108	-0,854	0,767	0,074	-0,185
d9	-0,002	0,501	-0,031	-0,850	0,778	0,019	-0,064
d10	-0,048	0,488	-0,084	-0,850	0,740	0,044	-0,128
d11	-0,334	0,554	-0,495	-1,000	0,571	0,057	-0,461
d12	-0,006	0,484	0,007	-0,737	0,826	0,031	0,036
d13	-0,028	0,510	-0,072	-0,834	0,830	0,052	-0,111
d14	-0,074	0,505	-0,134	-0,864	0,741	0,042	-0,204

Este tipo de análise não pode ser feito com o modelo PN e mostra as possíveis vantagens na interpretação psicométrica dos modelos TRI-NA propostos.

5.3.5 A probabilidade de resposta correta não condicional

Como Albert e Ghosh (2000) dizem, pode ser difícil interpretar o parâmetro de dificuldade b_j devido a que não pode ser expresso numa escala de probabilidade. Uma medida alternativa da dificuldade é a probabilidade de resposta correta não condicional (ver capítulo 4), isto é, a probabilidade de que um indivíduo escolhido aleatoriamente de uma população obtenha uma resposta correta para o j -ésimo item, também interpretada como a proporção esperada de alunos que podem responder corretamente o j -ésimo item. A Tabela 5.6 apresenta as probabilidades de resposta correta não condicionais sob os modelos PN e PAN e as proporções observadas de alunos na amostra, os quais respondem corretamente o j -ésimo ítem. A probabilidade não condicional é obtida usando a expressão no Apêndice 6 e é estimada usando a mediana a posteriori dos parâmetros.

Tabela 5.7 Comparação das proporções estimadas de resposta correta sob os modelos PN e PAN com as proporções observadas na amostra

Item	d estimado	Proporção observada	Proporção estimada		Diferenças estimada-observada	
			PN	PAN	PN	PAN
1	-0,084	0,794	0,786	0,789	0,0075	0,0052
2	-0,053	0,855	0,851	0,854	0,0042	0,0007
3	0,010	0,519	0,510	0,517	0,0092	0,0022
4	-0,175	0,931	0,918	0,923	0,0134	0,0078
5	-0,136	0,870	0,863	0,872	0,0075	-0,0014
6	0,027	0,366	0,361	0,361	0,0053	0,0055
7	-0,219	0,924	0,914	0,921	0,0102	0,0026
8	-0,108	0,878	0,860	0,870	0,0175	0,0077
9	-0,031	0,786	0,782	0,786	0,0044	0,0004
10	-0,084	0,863	0,854	0,861	0,0091	0,0015
11	-0,495	0,931	0,914	0,929	0,0173	0,0021
12	0,007	0,351	0,347	0,349	0,0036	0,0024
13	-0,072	0,824	0,818	0,823	0,0066	0,0010
14	-0,134	0,947	0,943	0,948	0,0040	-0,0012

A probabilidade não condicional de resposta correta sob os modelos PN e PAN pode ser considerada com a proporção estimada de resposta correta e pode ser comparada com a proporção observada da amostra. A comparação das diferenças (estimada menos observada) com o parâmetro de penalidade é mostrada na Figura 5.6. Os círculos pretos correspondem as diferenças sob o modelo PAN, e os círculos transparentes correspondem as diferenças sob o modelo PN. Note que a distância entre as diferenças na proporção de resposta correta sob os modelos PN e PAN são altas quando o parâmetro de penalidade é alto como no caso do item 11. Assim, existe uma forte evidência que para alguns itens do conjunto de dados, o modelo PAN é o mais apropriado.

5.4 Discussão

Os conceitos de tamanho de amostra efetivo (ESS) como definido em Kass *et al.* (1998) e o tamanho da amostra por segundo (ESS/s) como considerado em Sahu (2002) foram considerados para o estudo da eficiência na estimação dos parâmetros do modelo introduzido. Nós chamamos a atenção para a importância de pesquisar tais quantidades quando uma abordagem de dados aumentados é usada na família TRI-NA e existe uma forte evidência de autocorrelação nas

estimativas dos parâmetros que segue das cadeias geradas. Embora a ergodicidade da cadeia de Markov segue das condições de regularidade requeridas para a convergência das cadeias de Markov, a presença de altas autocorrelações leva a necessidade de um grande número de iterações para avaliar a variabilidade das estimativas (média amostral, moda amostral, quantiles). Para obter amostras com baixa autocorrelação sub amostragem sistemática das cadeias de Markov são sugeridas, isto também é uma ferramenta útil para minimizar requerimentos de armazenamento. Com os dados do teste de Matemática da Seção 5.2, não foi possível detectar assimetria quando uma priori difusa foi considerada para o parâmetro b . Entretanto, isto foi possível com prioris mais precisas. Subseqüentes pesquisas no estudo da sensibilidade da posteriori de d (ou λ) com respeito à especificação de priori diferentes são necessárias.

Outra contribuição do capítulo é a facilidade de implementar o modelo TRI-NA usando a versão da verossimilhança proposta no Capítulo 4 que é diferente de se considerar a ligação proibito assimétrica de Chen *et al.* (1999).

Outra contribuição do capítulo é o estudo sobre comparação de modelos. Todos os modelos da TRI-NA são analisados usando DIC (Spiegelhalter *et al.*, 2002), $EAIC$ e $EBIC$ (Carlin e Louis, 2000 and Brooks, 2002). Nós também introduzimos residuais latentes para os modelos e uma medida de discrepância global como é o caso da soma de quadrados residuais latentes a posteriori, o qual pode ser usado para comparação dos modelos TRI-NA de maneira conveniente.

Finalmente, do ponto de vista do pessoal que elabora itens, a presença de um novo parâmetro de item que pode explicar a conduta assimétrica da CCI em termos de variações da probabilidade de sucesso para diferentes níveis de habilidade pode ser usada no desenvolvimento de testes mais precisos na estimação das habilidades dos examinados.

Nós apresentamos algumas interpretações para o novo parâmetro de item proposto denominado parâmetro de penalização no contexto do conjunto de dados do teste de matemática considerado. Nós também mostramos que o modelo PAN resulta ser o que melhor ajusta os dados observados. Embora os intervalos HPD na Tabela 5.6 indiquem os parâmetros de penalização não são significativamente diferentes de zero, e portanto um modelo PN poderia ser mais adequado para ajustar os dados, nós preferimos basear nossa conclusão usando os critérios tais como DIC, AIC e BIC esperado e a comparação entre proporções de resposta estimada e observada. Isto devido ao fato de que estimativas dos parâmetros de penalização apresentam uma grande variância, o qual poderia ser explicado pelo tamanho pequeno da amostra de indivíduos.

Extensões para considerar um modelo TRI-NA multidimensional, modelos hierárquicos TRI-NA, modelos TRI-NA múltível podem ser propostos em futuras pesquisas. Adaptações dos algoritmos propostos no Capítulo 4 para estes casos podem ser feitas facilmente. Outras extensões da ligação proibito assimétrica para respostas ordinais como em Johnson (2003) podem ser também estendidas para correspondentes modelos TRI-NA de resposta ordinal.

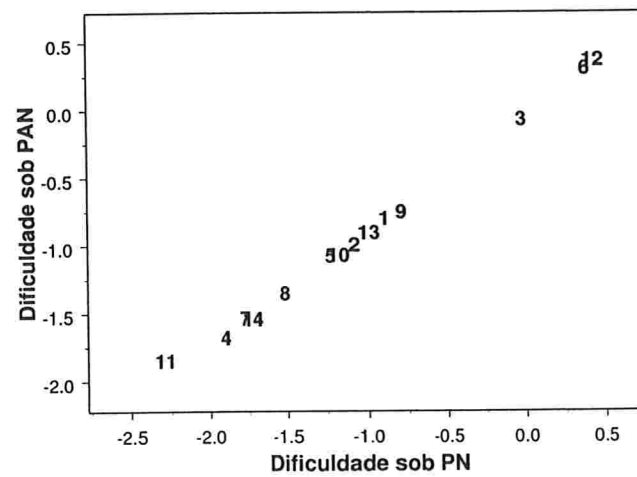
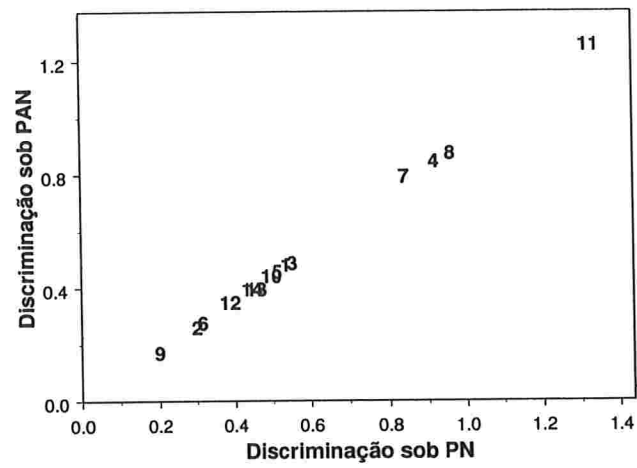


Figura 5.2 Parâmetros de discriminação e dificuldade sob os modelos PN e PAN

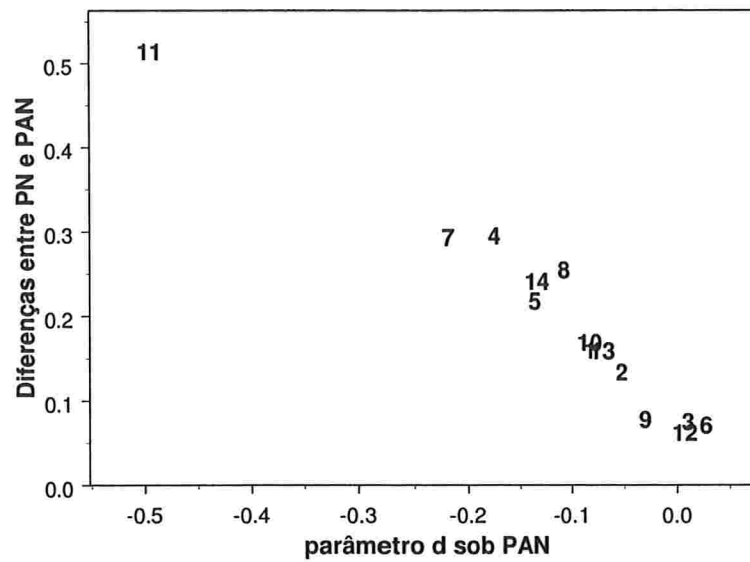


Figura 5.3 diferenças nos modelos PN e PAN $|\bar{a}| + |\bar{b}|$ de acordo as estimativas do parâmetro d dos itens

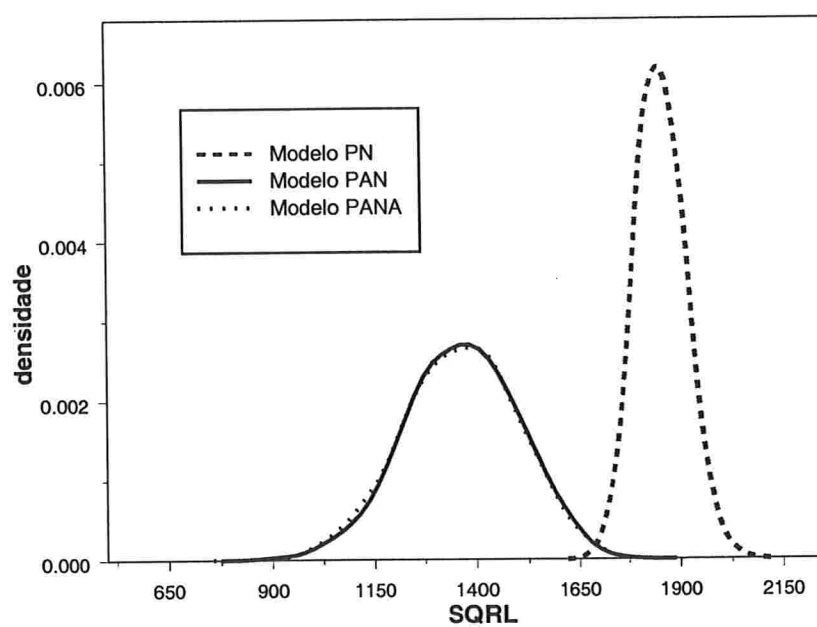


Figura 5.4 Soma de quadrados de residuais latentes (SSR) para os modelos PN, PAN e PANA

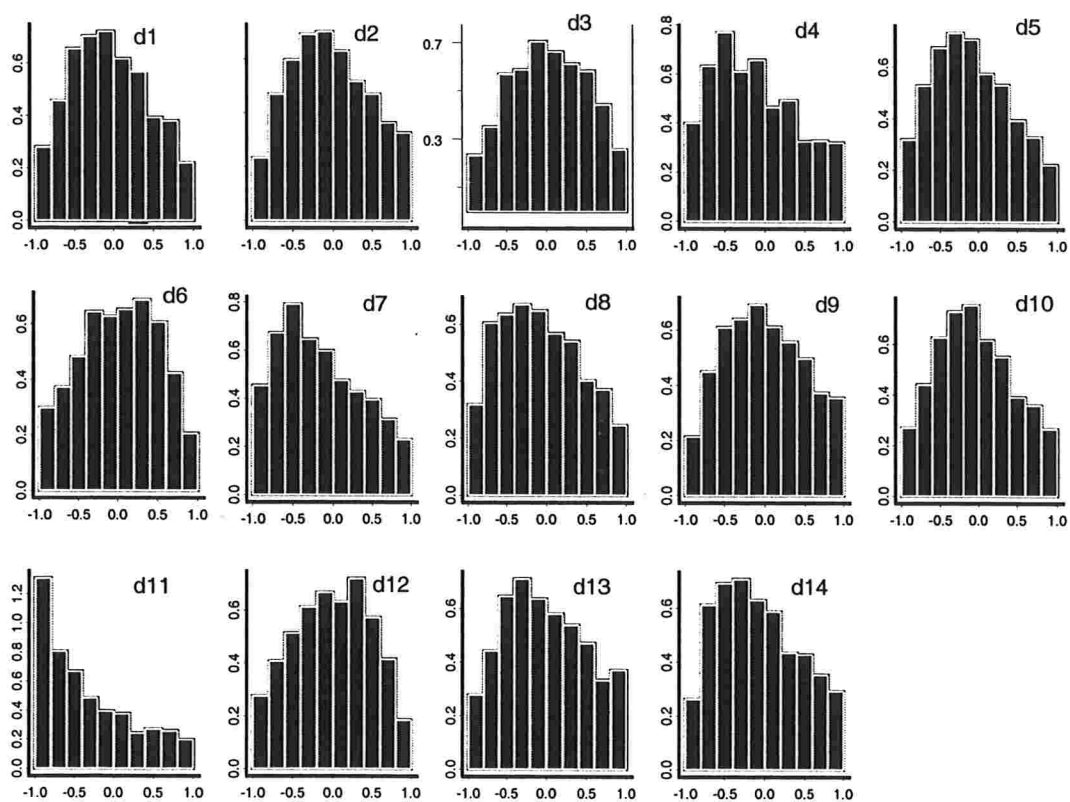


Figura 5.5 *Histogramas da distribuição posterior para o parâmetro d*

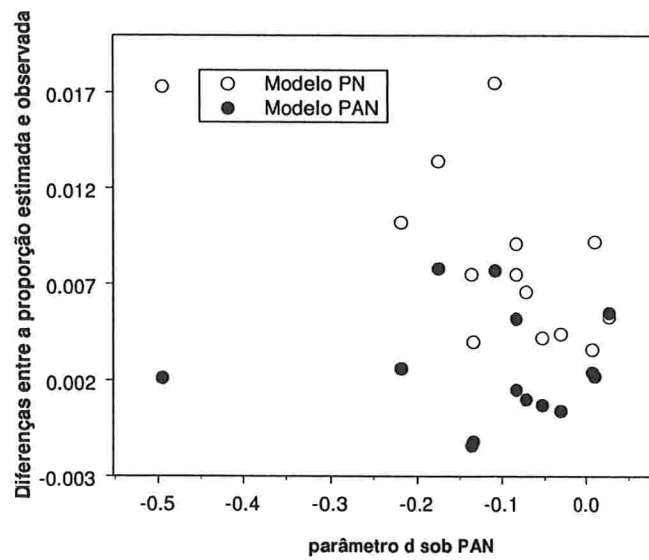


Figura 5.6 Comparação das diferenças (estimadas-observadas) sob os modelos PN e PAN para o parâmetro de penalidade d

Capítulo 6

Considerações Finais

O presente trabalho esteve motivado por dois grandes propósitos: I) Pelo desenvolvimento de um modelo de resposta ao item que se afaste da suposição de normalidade o que tem dominado a pesquisa e as aplicações destes modelos nos últimos 50 anos, e II) o desenvolvimento de modelos simples de ser implementados, estendidos, interpretados e aplicados sob a perspectiva Bayesiana e mediante o uso de programas livres.

A suposição de normalidade esta presente ao se assumir que as variáveis latentes - habilidades ou proficiência na avaliação escolar - são “bem comportadas” de acordo à normalidade considerando um comportamento idealizado das variáveis para descrever a conduta humana, semelhante à inteligência, como fora inicialmente proposto por Quetelet no século 19. A suposição de normalidade esta presente também no uso de ligações simétricas como o probito (também do logito, que apenas é uma metodologia aproximada) para as curvas características dos ítems ou funções de resposta dos ítems que modelam a probabilidade de resposta correta em função da variável latente e de parâmetros associados aos ítems. Embora, a idéia de afastar-se de tais suposições tenha sido proposta por Samejima (1997, 2000) nós desconhecemos até o momento, uma aplicação de outro modelo TRI assimétrico e menos ainda sob uma perspectiva Bayesiana.

Os métodos freqüentistas tem dominado as aplicações na TRI nos últimos 20 anos a partir do trabalho de Bock e Aitken (1981) cuja metodologia de estimação tem sido estendida para quase todos os modelos paramétricos propostos na TRI e tem sido difundida pelo uso de programas comerciais entre eles o BILOG embora ele também sea usado para uma aproximação bayesiana sem MCMC. No entanto diversas limitações tem sido indicadas na literatura, entre elas Kim *et al.* (1994), Patz e Junker (1999) e Sahu (2002). Além disso, o uso das soluções na TRI é particularmente dominada por especialistas e está generalizada nos grandes sistemas de avaliação educacional, não sendo freqüente sua implementação por profissionais da educação e pesquisadores mais interessados nas aplicações como psicólogos como pode ser notado nas publicações destas áreas.

Para contornar estes problemas nós nos propusemos desenvolver um novo modelo da TRI

para resposta binária que se afastasse da suposição de normalidade mas que tivesse como caso particular o modelo já difundido. Assim como, desenvolver uma metodologia de estimação sob a perspectiva bayesiana que seja simples de ser implementada, estendida, interpretada e aplicada sob uso de programas livres. Isto nos levou ao desenvolvimento da família normal assimétrica da Teoria da resposta ao item (TRI-NA) proposta neste trabalho e ao uso do programa WinBUGS.

É importante dizer que o estudo das propriedades da família TRI-NA levou ao estudo das propriedades correspondentes no modelo Probito-Normal e não o inverso, percebendo as vezes que nem sempre estes estudos foram feitos. Esta revisão envolvendo estudo de propriedades do modelo Probito-Normal são apresentada no Capítulo 2.

Por outro lado, no momento de implementar o uso da família TRI-NA algumas dúvidas surgiram ao respeito da nova ligação proposta. Isto aconteceu porque esta proposta não seguiu o desenvolvimento tradicional das ligações binárias usadas na TRI. Isto é, na ordem histórica, primeiro é proposto uma nova ligação para resposta binária em regressão e após esta é usada em modelos mais complexos como na TRI.

Assim, embora a ordem histórica possa ser completamente diferente, na tese são apresentados os três tipos de desenvolvimentos deste trabalho no ordem que achamos adequada: a) O Modelo Probito-Normal na Teoria da Resposta ao Item, b) Uma função de ligação probito assimétrica generalizada para resposta binária e c) A família normal assimétrica da Teoria da resposta ao item.

6.1 Conclusões

Em relação ao Capítulo 3 nós mostramos que

- Modelos probito assimétrico generalizados são mais flexíveis para o modelagem de resposta binária que precisam ligações assimétricas.
- A assimetria da ligação esta associada aos parâmetros da distribuição acumulada usada na definição da probabilidade e o parâmetro de assimetria é independente do intervalo de variação do preditor linear.
- Trata-se de uma função de ligação que apresenta o caso simétrico como caso particular a ser determinado pelos dados.
- Existem diversas versões da verossimilhança que facilitam o uso de diferentes métodos de estimação incluindo abordagem freqüentista e bayesiana.
- Proposições garantem a existência de estimativas de máxima verossimilhança e a da distribuição posterior quando prioris impróprias são usadas tanto na abordagem condicional e não condicional.

- Considerando o resultado obtido na proposição A.1 no Apêndice para a fda da distribuição normal assimétrica como a avaliação de uma certa normal bivariada para valores condicionados de uma outra variável de resposta positiva o uso de ligação proibito assimétrica generalizada pode ser justificada para uma resposta binária condicionada numa outra variável de resposta positiva. Por exemplo poupança ou não para determinado nível de renda.

Em relação ao Capítulo 4 nós mostramos que

- A nova curva característica do item incorpora um parâmetro (de assimetria) que tem interpretação no contexto do modelo e dá maior flexibilidade a CCI
- Diversas versões da verossimilhança e de verossimilhança aumentada são disponíveis para a TRI-NA sendo possível a escolha mais pertinente com fins computacionais.
- O afastamento da suposição de normalidade das variáveis latentes pode levar ao uso de modelos TRI para variáveis até agora afastadas do uso da TRI (mais ligada a variáveis cognitivas e educacionais) como variáveis clínicas, afetivas e emocionais, variáveis nas que é esperada uma distribuição assimétrica teórica como acontece por exemplo com a depressão.
- Os parâmetros tradicionais de item, discriminação e dificuldade são igualmente interpretáveis em qualquer dos modelos TRI-NA.
- Os modelos TRI-NA que incorporam assimetria podem ser mais convenientes para justificar a verdadeira chance de acertar um ítem quando algum tipo de assimetria pode ser justificada, mas quando não é o caso recuperam o modelo anterior. Assim temos um modelo mais geral e único para ambas situações.
- O modelo TRI-NA pode ser aplicado em situações de populações restritas em que existe certa evidência de assimetria devida à restrição da população de interesse. Por exemplo em populações carentes ou em populações favorecidas, em populações bilingües, comunidades nativas, regiões, na posgraduação, etc. Estudos posteriores podem ser ilustrativos deste uso de suas vantagens em relação a modelos TRI simétricos.

Finalmente, os programas desenvolvidos para este trabalho constituem uma “library” de modelos que podem ser aplicados facilmente usando um programa livre por usuários e pode permitir o desenvolvimento de extensões ou modificações do modelo por pesquisadores.

6.2 Futuros Trabalhos

O presente trabalho, constitui um primeiro estágio na tentativa de se escrever um livro de Inferência Bayesiana na TRI ausente não somente na língua portuguesa mais também na literatura internacional. Um livro que possa combinar argumentos teóricos para o desenvolvimento

da pesquisa nesta área mas também metodologias fáceis de serem implementadas em programas livres.

Em relação ao Capítulo 3 nós propomos os seguintes trabalhos

- Generalização do modelo de regressão binária usando a ligação probito assimétrica generalizada para modelos mistos e modelos com erros de medição.
- Aplicação do modelo de regressão binária usando a ligação probito assimétrica generalizada para modelos binomiais, estudos epidemiológicos, modelagem multinível, análise de dados longitudinais, meta-análise e teoria da resposta ao item.
- Estender o uso da ligação proposta para dados assimétricos comuns nas ciências sociais, ciências médicas, agricultura, genética e ciências do comportamento.
- Implementação da solução frequentista para ser introduzida como procedimento padrão em programas como Splus e SAS.
- Elicitação de prioris por especialistas (conhecedores da área da aplicação) para p_i e não para β , como proposto em Bedrick *et al.* (1996) ou elicitação de prioris usando dados históricos como proposto em Chen, *et al.* (2001).
- Desenvolver extensões da ligação probito assimétrica generalizada para resposta ordinais como em Johnson (2003).

Em relação ao Capítulo 4 nós propomos os seguintes trabalhos:

- É possível desenvolver importantes extensões da TRI-NA para modelos que incorporem o parâmetro de adivinhação ou "chute" como acontece com o modelo PN (veja-se Sahu, 2002; Glass e Meijer, 2003), Modelo multidimensionais TRI-NA (veja-se Beguin e Glass, 2001; Simon, 2001; Linardakis e Dellaportas, 2002 para o modelo PN), Modelo multinível (veja-se Fox e Glass, 2001 para o modelo PN), Teslet (veja-se Wang *et al.*, 2002 para o modelo PN), Análise fatorial confirmatória (veja-se Segall, 2003 para o modelo PN) e Modelo com erro de medida (Fox e Glass, 2003 para o modelo PN)
- Desenvolvimento de modelos hierárquicos em todas as extensões propostas acima.
- Análise da sensibilidade a escolha de prioris diferentes e ao desenvolvimento de critérios de monitoramento da convergência multidimensionais para os diferentes parâmetros envolvidos na TRI-NA.
- Tomando em conta que diversas versões da verossimilhança e verossimilhança para dados aumentados são disponíveis, pode-se sugerir uma análise de eficiência da estimação dos parâmetros sob estas diferentes versões.

-
- Desenvolver e implementar outros procedimentos de comparação de modelos para a escolha do melhor modelo que se ajusta a um particular conjunto de dados usando os diversos modelos apresentados que sejam próprios na TRI.
 - Desenvolver análise de ajuste e avaliação de observações influentes baseados em residuais latentes propostos.
 - Desenvolver uma metodologia mais afinada para a estimação das habilidades latentes para estudos de grande escala como os do ENEM, SAEB e demais sistemas de avaliação.
 - Um ponto interessante seria desenvolver estudos de simulação para investigar a eficiência de estimações nos Modelos TRI-NA.

Apêndice A

A distribuição normal assimétrica

Na tese nós usamos a normal assimétrica, a normal assimétrica padrão devida a Azzalini (1985) e a normal assimétrica devida a Chen *et al.* (1999). Propriedades para ditas distribuições são dadas em diversos artigos, entre eles, Azzalini (1985), Henze (1986), Arnold *et al.* (1993), Azzalini e Dalla Valle (1996), Loperfido (2002), Liseo e Loperfido (2003) e Dalla Valle (2004), não em tanto algumas propriedades importantes usadas na tese são listadas aqui.

Uma variável R segue uma distribuição normal assimétrica (Dalla Valle, 2004), com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$ em que $\mu \in \mathbb{R}$ é um parâmetro de locação e $\sigma^2 > 0$ é um parâmetro de escala e $\lambda \in \mathbb{R}$ é um parâmetro de assimetria, se a função de densidade de R é dada por

$$f_{\theta}(r) = \frac{2}{\sigma} \phi\left(\frac{r - \mu}{\sigma}\right) \Phi\left(\lambda \frac{r - \mu}{\sigma}\right), \quad (\text{A.1})$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ denotam, respectivamente, a função de densidade e de distribuição da distribuição normal padrão. A notação considerada é $R \sim NA(\theta)$ onde $\theta = (\mu, \sigma^2, \lambda)$, mas também se pode usar a notação $\phi_{NA}(r; \mu, \sigma^2, \lambda)$ para representar a densidade. $F_{\theta}(r)$, com $\theta = (\mu, \sigma^2, \lambda)$, denota a fda da distribuição normal assimétrica. Quando considerada a transformação $d = \frac{\lambda}{\sqrt{1+\lambda^2}}$ a distribuição normal assimétrica tem outra parametrização em que $-1 \leq d \leq 1$. Para voltar à parametrização original considere $\lambda = \frac{d}{(1-d^2)^{1/2}}$.

Casos particulares da normal assimétrica

- Se $\lambda = 0$, i.e se $\theta = (\mu, \sigma^2, 0)$ a densidade de R em (A.1) se reduz a densidade da $N(\mu, \sigma^2)$. Em particular se $\theta = (0, 1, 0)$ obtemos a normal padrão.
- No caso especial de $\theta = (0, 1, \lambda)$, a densidade de R em (A.1) se reduz a densidade da normal assimétrica padrão dada por Azzalini (1985) e neste caso escrevemos que $S \sim NA(\lambda)$ cuja função de densidade é dada por:

$$f_{\lambda}(s) = 2\phi(s)\Phi(\lambda s) \quad (\text{A.2})$$

Em particular se para a normal assimétrica padrão fazemos $\lambda \rightarrow \infty$ nos obtemos a densidade da *Half Normal* o normal positiva (Chen *et al.*, 1999) e escrevemos $V \sim HN(0, 1)$ e

sua função de densidade é dada por

$$g(v) = \frac{2}{\sqrt{2\pi}} e^{-v^2/2} \quad (\text{A.3})$$

Em que $E(v) = \frac{2}{\pi}^{1/2}$, $V(v) = 1 - \frac{2}{\pi}$.

- No caso especial de $\theta = (0, 1, -\lambda)$, a densidade de R em (A.1) se reduz a densidade da normal assimétrica padrão com parâmetro $-\lambda$ a qual é chamada aqui a normal assimétrica *BBB* e neste caso escrevemos que $S \sim NA_{BBB}(\lambda)$ cuja função de densidade é dada por:

$$f_\lambda(s) = 2\phi(s)\Phi(-\lambda s) \quad (\text{A.4})$$

- No caso especial de $\theta = (0, 1 + \lambda^2, \lambda)$, a densidade de R em (A.1) se reduz a densidade da normal assimétrica padrão dada por Chen *et al* (1999) e Branco e Dey (2002) a qual por sua vez é um caso particular da normal assimétrica dada em Sahu *et al.* (2003), e neste caso escrevemos que $S \sim NA_{CDS}(\lambda)$ chamada aqui normal assimétrica *CDS* cuja função de densidade é dada por:

$$f_\lambda(s) = \frac{2}{\sqrt{1+\lambda^2}} \phi\left(\frac{s}{\sqrt{1+\lambda^2}}\right) \Phi\left(\lambda \frac{s}{\sqrt{1+\lambda^2}}\right), \quad (\text{A.5})$$

Propriedades úteis da normal acumulada

As seguintes propriedades são úteis para o cálculo usando distribuições normais assimétricas porque aparecem com freqüência.

- Se Z é v.a $\sim N(0, 1)$, temos que $E(\Phi(hZ + k)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right)$ (Azzalini, 1985)
- Se $Z \sim N_k(\mathbf{0}, \Sigma)$, temos que, $E_Z[\Phi(\mathbf{u} + \mathbf{a}'Z)] = \Phi\left(\frac{\mathbf{u}}{(1+\mathbf{a}'\Sigma\mathbf{a})^{1/2}}\right)$ (Azzalini e Dalla Valle, 1996).
- Se $Z \sim N_k(\mathbf{0}, I_k)$, temos que, $E_Z[\phi_m(\mathbf{u} + AZ; \mathbf{0}, \Omega)] = \phi_m(\mathbf{u}; \mathbf{0}, \Omega + AA')$ e $E_Z[\Phi_m(\mathbf{u} + AZ; \mathbf{0}, \Omega)] = \Phi_m(\mathbf{u}; \mathbf{0}, \Omega + AA')$ (Arellano e Genton, 2003).
- Se $Z \sim N_k(\boldsymbol{\mu}_1, \Sigma)$, e \mathbf{u} um vetor m dimensional e seja \mathbf{A} uma matriz de posto completo, com $m \leq k$, temos que,
 $E_Z[\Phi_m(\mathbf{u} + AZ; \boldsymbol{\mu}_2, \Omega)] = \Phi_m(\mathbf{u}; \boldsymbol{\mu}_2 - \mathbf{A}\boldsymbol{\mu}_1; \Omega + \mathbf{A}\Sigma\mathbf{A}')$ (Liseo e Loperfido, 2003).

Propriedades úteis da normal assimétrica padrão de Azzalini

$S \sim NA(\lambda)$ é a distribuição normal assimétrica padrão, sua função de densidade dada em (A.2) pode ser denotada também por $\phi_{NA}(s; 0, 1; \lambda)$ ou simplesmente por $\phi(s; \lambda)$. Como observado em Azzalini (1985) temos que:

1. $-S$ é $NA(-\lambda)$ que corresponde à normal assimétrica *BBB*,
2. A densidade de S é fortemente unimodal, i.e $\log\phi(s; \lambda)$ é função côncava de s .

A função de distribuição acumulada de S pode ser denotada por $\Phi(s; \lambda)$ a qual é dada em Azzalini (1985) por

$$\Phi(s; \lambda) = 2 \int_{-\infty}^s \int_{-\infty}^{dm} \phi(m)\phi(n)dn dm$$

que pode ser calculada usando a função $T(h; a)$ estudada por Owen (1956), que da uma integral da densidade normal bivariada padrão sobre a região limitada pelas linhas $x = h, y = 0$, e $y = ax$ no plano (x, y) . Então obtemos

$$\Phi(s; \lambda) = \Phi(s) - 2T(s; \lambda)$$

em que

$$T(s; \lambda) = \int_s^{\infty} \int_0^{dm} \phi(m)\phi(n)dn dm.$$

A $T(s; \lambda)$ é uma função decrescente de h que satisfaz as seguintes propriedades:

$$T(-h; \lambda) = T(h; \lambda), \quad -T(h; \lambda) = T(h; -\lambda), \quad 2T(h; 1) = \Phi(h)\Phi(-h)$$

as quais levam as seguintes propriedades adicionais dadas por Azzalini (1985)

3. $1 - \Phi(-s; \lambda) = \Phi(s; -\lambda)$
4. $\Phi(s; 1) = \Phi(s)^2$
5. $\sup |\Phi(s) - \Phi(s; \lambda)| = \frac{1}{\pi} \arctan |\lambda|$

Representações da normal assimétrica padrão de Azzalini

Embora simular da normal assimétrica padrão seja simples devido as diversas bibliotecas que podem ser obtidas da WWW page: <http://azzalini.stat.unipd.it/SNw>, as seguintes representações são úteis para entender a relação da normal assimétrica com a normal.

- Seja $W \sim N(0, 1)$ e $V \sim HN(0, 1)$ v.a independentes e $|d| < 1$. Temos que a distribuição de

$$S = dV + (1 - d^2)^{1/2}W$$

é $NA(\lambda)$. Esta é a representação estocástica marginal e foi provada por Henze (1986), Azzalini e Dalla Valle (1996).

- Se $V, W \sim N(0, 1)$ são v.a independentes, a distribuição de $S = W|V < \lambda W$ é $NA(\lambda)$. Veja Henze (1986), Azzalini e Dalla Valle (1996)

- Se $S|V \sim N(dv, 1 - d^2)$ e $V \sim HN(0, 1)$ (densidade normal positiva), a distribuição de S é $NA(\lambda)$.
- Seja $\begin{pmatrix} V \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & d \\ d & 1 \end{pmatrix}\right)$ e $|d| < 1$. Temos que $W|V > 0 \sim NA(\lambda)$. Esta é a representação estocástica condicional e foi provada por Arnold *et al* (1993) e Azzalini e Dalla Valle (1996).
- Seja $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & d \\ d & 1 \end{pmatrix}\right)$ e $|d| < 1$. Temos que

$$S = \begin{cases} W & \text{Se } V > 0, \\ -W & \text{c.c.} \end{cases} \text{ é } NA(\lambda).$$
 Esta outra representação estocástica condicional foi provada por Azzalini e Capitanio (1999).
- Seja $\begin{pmatrix} V \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. Temos que

$$\max(V, W) \sim NA(\lambda) \text{ e } \min(V, W) \sim NA(-\lambda) \text{ com } \rho = \frac{1-\lambda^2}{(1+\lambda^2)} \text{ e } \lambda = \sqrt{\frac{1-\rho}{1+\rho}},$$
 Esta representação foi provada por Loperfido (2002).

Propriedades da normal assimétrica padrão CDS

$S \sim NA_{CDS}(\lambda)$ é a distribuição normal assimétrica padrão CDS, e sua função de densidade dada em (A.3) pode ser denotada também por $\phi_{CDS}(s; 0, 1; \lambda)$ ou simplesmente por $\phi_{CDS}(s; \lambda)$. Como observado em Chen *et al.* (1999) temos as seguintes representações são úteis para entender a relação da normal assimétrica CDS com a normal.

- Seja $W \sim N(0, 1)$ e $V \sim HN(0, 1)$ v.a independentes e $\lambda \in \mathbb{R}$. Temos que a distribuição de $S = \lambda V + W$ é $NA_{CDS}(\lambda)$.
- Se $S|V \sim N(\lambda v, 1)$ e $V \sim HN(0, 1)$, a distribuição de S é $NA_{CDS}(\lambda)$.

Propriedades da normal assimétrica

Dado R que segue uma distribuição normal assimétrica com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$, i.e $R \sim NA(\mu, \sigma^2, \lambda)$

a) A função geradora de momentos é dada por

$$M_R(t) = E(e^{tR}) = 2 \exp\left(\mu t + \frac{1}{2} \sigma^2 t^2\right) \Phi(\sigma t)$$

b) Em particular a média e variância são dadas por:

$$\mu_R = E[R] = \mu + \sqrt{\frac{2}{\pi}}d\sigma = \mu + 0.79787846d\sigma;$$

$$\sigma_R^2 = Var[R] = (1 - \frac{2}{\pi}d^2)\sigma^2 = (1 - 0.6366198d^2)\sigma^2;$$

Aproximando até 2 decimais, note que $\mu_R \in [\mu - 0.8\sigma, \mu + 0.8\sigma]$ e qual cresce conforme $|d|$ cresce e $\sigma_R^2 \in [0.36\sigma^2, \sigma^2]$ o qual decresce conforme $|d|$ cresce.

c) A Assimetria γ o tercer momento é dada por $\gamma = \mu_3^2 = E(\frac{R-\mu_R}{\sigma_R})^3$

d) A Curtosis κ o cuarto momento é dada por $\kappa = \mu_4^2 = E(\frac{R-\mu_R}{\sigma_R})^4$

e) Em particular nós obtemos que

	<i>NA CDS</i>	<i>NA Azzalinni</i>
$M_S(t)$	$2exp(\frac{1}{2}t^2) \Phi(dt)$	$2exp(\frac{1}{2}(1 + \lambda^2)t^2) \Phi(\lambda t)$
μ_S	$(\frac{2}{\pi})^{1/2}\lambda$ 0.79787846λ	$(\frac{2}{\pi})^{1/2}d$ $0.7978846d$
σ_S^2	$1 + (1 - \frac{2}{\pi})\lambda^2$ $1 + 0.3633802\lambda^2$	$1 - \frac{2}{\pi}d^2$ $1 - 0.3633802d^2$
γ	$(\frac{2}{\pi})^{3/2}(1 - \frac{2}{\pi})sig(\lambda) \frac{\lambda^3}{(1+(1-\frac{2}{\pi})\lambda^2)^{3/2}}$ $(\frac{2}{\pi})^{3/2}(1 - \frac{2}{\pi})sig(d) \frac{d^3}{(1-\frac{2}{\pi}d^2)^{3/2}}$ $-0.843 \leq \frac{0.1845787d^3}{(1-0.6366198d^2)^{3/2}} \leq 0.843$	$(\frac{2}{\pi})^{3/2}(2 - \frac{\pi}{2})sig(\lambda) \frac{\lambda^3}{(1+(1-\frac{2}{\pi})\lambda^2)^{3/2}}$ $(\frac{2}{\pi})^{3/2}(2 - \frac{\pi}{2})sig(d) \frac{d^3}{(1-\frac{2}{\pi}d^2)^{3/2}}$ $0.995 \leq \frac{0.2180136d^3}{(1-0.6366198d^2)^{3/2}} \leq 0.995$

em que $sign(\cdot)$ é a função sinal que toma valor 1 quando seu argumento é positivo e valor -1 em outro caso.

- Na normal assimétrica *CDS*, aproximando até 2 decimais, temos que $\mu_S \in [-0.80|\lambda|, 0.80|\lambda|]$, $\sigma_S^2 \in [1, 1 + 0.36\lambda^2]$ e $\gamma \in [-0.84, 0.84]$. Note que $\mu_S \in \mathbb{R}$, $\sigma_S^2 \geq 1$ e que γ é baixo.
- Na normal assimétrica de Azzalinni temos que $\mu_S \in [-0.80, 0.80]$, $\sigma_S^2 \in [0.36, 1]$ e $\gamma \in [-1.00, 1.00]$. Note que tanto μ_S , σ^2 e γ são baixos.

f) A relação que existe entre a normal assimétrica padrão devida a Azzalini e a normal assimétrica padrão *CDS* é a seguinte:

$$\text{Dada a v.a } S \sim NA(\lambda) \text{ então a v.a } S^* = \sqrt{1 + \lambda^2}S \sim NA_{CDS}(\lambda)$$

g) A variável $S = (R - \mu)/\sigma$ é distribuída de acordo com a distribuição normal assimétrica padrão.

h) A variável $R^* = a + bR \sim NA(a + b\mu, b^2\sigma^2, sign(b)\lambda)$ em que $sig(\cdot)$ é a função signo que toma valor 1 se seu argumento é positivo e -1 caso contrário. Em particular dado $S \sim NA(\lambda)$, então $S^* = a + bS \sim NA(a, b^2, sign(b)\lambda)$ e por tanto $R = \mu + \sigma S \sim NA(\mu, \sigma^2, \lambda)$ e $R = \mu - \sigma S \sim NA(\mu, \sigma^2, -\lambda)$.

i) O índice de curtose da normal assimétrica de Azzalini é dada por:

$$\kappa = \frac{8(\pi - 3)d^4}{(\pi - 2d^2)^2}$$

implicando $0 < \kappa < 0.8692$.

j) Quando fazemos $\mu = -(\frac{d^2}{\pi - d^2})^{1/2} = -(\frac{2d^2}{\pi - 2d^2})^{1/2}$ e $\sigma^2 = \frac{1}{1 - \frac{2d^2}{\pi}} = \frac{\pi}{\pi - 2d^2}$ em (A.1) nós temos que $\mu_R = 0$ e $\sigma_R^2 = 1$ e neste caso temos uma interessante distribuição normal assimétrica padrão.

Proposição A.1

E fácil mostrar que a fda da distribuição normal assimétrica é dada por

$$F_\theta(r) = 2\Phi_2(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{A.6}$$

em que $\mathbf{x} = (r, 0)$, onde $\Phi_2(\cdot)$ denota a fda da normal bivariada padrão, com parâmetros $\boldsymbol{\mu} = (\mu, 0)$ e $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & -d \\ -d & 1 \end{pmatrix}$, em que $-d \in [-1, 1]$ é o coeficiente de correlação dado por $d = \frac{\lambda}{\sqrt{1 + \lambda^2}}$.

Prova. Dado $S \sim NA(0, 1, \lambda)$ temos que $F_\theta(s) = \int_{-\infty}^s 2\phi(t)\Phi(\lambda t)dt = \int_{-\infty}^s 2\phi(t)\Phi(0 - (-\lambda)t)dt = \int_{-\infty}^s 2\phi(t)\Phi(0 - \frac{-d}{\sqrt{1-d^2}})dt$, e considerando o resultado de Parrish e Bergmann (1981), o qual considera que a distribuição normal padrão bivariada com correlação ρ avaliada no ponto (h, k) pode ser escrita como $\Phi_2(h, k; \rho) = \int_{-\infty}^h \phi(w)\Phi(\frac{k-\rho w}{\sqrt{1-\rho^2}})$, tem-se que a acumulada da distribuição normal assimétrica padrão é $F_\theta(s) = \int_{-\infty}^s 2\phi(t)\Phi(\lambda t)dt = 2\Phi_2(s, 0; -\frac{\lambda}{\sqrt{1+\lambda^2}}) = 2\Phi_2(s, 0; -d)$. Finalmente note que $R = \sigma S + \mu$ e por tanto $F_\theta(r) = 2\Phi_2(\frac{r-\mu}{\sigma}, 0, -d)$ e de aqui se obtém o resultado desejado. \square

Observação 1. Este resultado indica que a fda da normal assimétrica avaliada no ponto r também é obtida considerando uma particular distribuição normal bivariada avaliada no ponto $(r, 0)$. Este resultado é importante pois vários algoritmos eficientes são disponíveis para o cálculo de integrais relativas à distribuição normal bivariada (veja Genz 1992, 1993). Outro algoritmo para a fda da normal assimétrica está baseado na função de Owen (Azzalini, 1985 e Dalla Valle, 2004) indicado acima o qual é disponível para os programas R e Matlab. Entretanto a distribuição de R é obtida considerando $(U, W) \sim N_2\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & -d \\ -d & 1 \end{pmatrix}\right)$ e definindo $R = U|W > 0$ (veja Arnold *et al.*, 2002), os quais definem uma interessante interpretação de fdp de R considerando truncamento seletivo (Greene, 2003) ou reporte seletivo (Arnold and Beaver, 2002).

No caso da distribuição normal assimétrica padrão se tem a seguinte expressão de seu fda

$$F_\lambda(s) = 2\Phi_2\left(\begin{pmatrix} s \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -d \\ -d & 1 \end{pmatrix}\right) \quad (\text{A.7})$$

A função de distribuição acumulada da normal assimétrica padrão denotada também por $\Phi_{SN}(z; \lambda)$ o simplesmente por $\Phi(z; \lambda)$ corresponde à função de distribuição de uma normal bivariada truncada padrão com correlação $-d$ (Arnold *et al.*, 1993), que podemos denotar por simplicidade por $\Phi_2[(s, 0)'; -d]$. Veja Arellano e Genton (2003) para uma expressão similar no caso da normal assimétrica multivariada.

No caso da distribuição normal assimétrica padrão *CDS* se tem a seguinte expressão de seu fda

$$F_\lambda^{CDS}(s) = 2\Phi_2\left(\begin{pmatrix} s \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{1-d^2} & -d \\ -d & 1 \end{pmatrix}\right) \quad (\text{A.8})$$

Também o seguinte resultado pode ser conveniente.

Proposição A.2

É fácil mostrar que a fda da normal assimétrica também é dada por

$$F_{\theta}(r) = \int_0^{\infty} g(v) \Phi\left(\frac{r - \mu - \sigma dv}{\sigma \sqrt{1 - d^2}}\right) dv = \int_0^{\infty} g(v) \Phi\left(\frac{r - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v\right) dv, \quad (\text{A.9})$$

onde $g(\cdot)$ é a fdp da Half Normal ou Normal positiva, $\Phi(\cdot)$ e $\phi(\cdot)$ denotam, respectivamente, a fda e fdp da distribuição normal padrão.

Prova

Considerando a representação condicional de $S \sim NA(\lambda)$ temos que $S|V \sim N(dv, 1 - d^2)$ e $V \sim HN(0, 1)$ e de aqui a função de densidade de S pode ser obtida integrando sobre o domínio de V a distribuição conjunta $(S|V, V)$ isto é: $f_{\lambda}(s) = \int_0^{\infty} \phi\left(\frac{s - dv}{\sqrt{1 - d^2}}\right) g(v) dv$ e por tanto a fda de S é

$$F_{\lambda}(s) = \int_{-\infty}^s \int_0^{\infty} \phi\left(\frac{t - dv}{\sqrt{1 - d^2}}\right) g(v) dv dt = \int_0^{\infty} \left[\int_{-\infty}^s \phi\left(\frac{t - dv}{\sqrt{1 - d^2}}\right) dt \right] g(v) dv$$

$$F_{\lambda}(s) = \int_0^{\infty} \Phi\left(\frac{s - dv}{\sqrt{1 - d^2}}\right) g(v) dv$$

Agora defina $R = \mu + \sigma S$ e por tanto $F_{\theta}(r) = P(R \leq r) = P(\mu + \sigma S \leq r) = P(S \leq \frac{r - \mu}{\sigma})$. De aqui temos que

$$F_{\theta}(r) = \int_0^{\infty} \Phi\left(\frac{r - \mu}{\sigma} - dv\right) 2\phi(v) dv$$

$$F_{\theta}(r) = \int_0^{\infty} \Phi\left(\frac{r - \mu - \sigma dv}{\sigma \sqrt{1 - d^2}}\right) g(v) dv = \int_0^{\infty} \Phi\left(\left(\frac{r - \mu}{\sigma} \sqrt{1 + \lambda^2} - \lambda v\right)\right) g(v) dv, \quad (\text{A.10})$$

No caso da distribuição normal assimétrica padrão tem-se a seguinte expressão de sua fda $F_{\lambda}(s) = \int_0^{\infty} g(v) \Phi\left(r \sqrt{1 + \lambda^2} - \lambda v\right) dv$ e no caso da distribuição normal assimétrica padrão *CDS* tem-se a seguinte expressão de sua fda $F_{\lambda}^{CDS}(s) = \int_0^{\infty} g(v) \Phi\left(s - \lambda v\right) dv$.

Proposição A.3

Note que podemos escrever $p = F_{\theta}(r)$ usando a seguinte estrutura lineal latente:

- $Y = I(T > 0) = \begin{cases} 1, & T > 0; \\ 0, & T \leq 0, \end{cases}$
- $T \sim NA(r - \mu, \sigma^2, -\lambda)$.

Prova

Considere $T = r + e$, $e \sim NA(-\mu, \sigma^2, -\lambda)$, então $p = P(Y = 1) = P(T > 0) = P(e + r > 0) = P(e > -r) = 1 - P(e \leq -r) = 1 - P\left(\frac{e + \mu}{\sigma} \leq \frac{-r + \mu}{\sigma}\right) = 1 - \Phi_{NA}\left(\frac{-r + \mu}{\sigma}, -\lambda\right)$ A qual é igual

a $\Phi_{NA}(\frac{r-\mu}{\sigma}, \lambda)$ considerando a propriedade 3) acima. Finalmente temos que $p = \Phi_{NA}(\frac{r-\mu}{\sigma}, \lambda) = F_{\theta}(r)$ em que $\theta = (\mu, \sigma^2, \lambda)$.

Apêndice B

Critérios de comparação de modelos no contexto Bayesiano

O desenvolvimento dos métodos MCMC tem possibilitado ajustar uma grande quantidade de modelos para um mesmo conjunto de dados, mas isto leva a la necessidade de comparar modelos alternativos com o propósito de identificar um modelo sucinto que possa descrever adequadamente estes dados. Existe uma variedade de metodologias para comparar o ajuste de modelos Bayesianos alternativos. Veja por exemplo o trabalho de Gilks *et al.* (1996). O critério principal usado neste trabalho é o *Critério de Informação de Desvio* (Deviance Information Criterion) (*DIC*) proposto por Spiegelhalter *et al.* (2002). O *DIC* é útil aqui porque é relativamente fácil de calcular, interpretar e é bem sucedido para modelos hierárquicos complicados (Johnson, 2003) e esta baseada na noção de “Desvio Bayesiano” (Dempster, 1977).

O *desvio bayesiano* (Dempster, 1977) para os modelos TRI-NA nos capítulos 2,4 e 5 e para os modelos de regressão proposto no capítulo 3, é definido, de maneira geral, usando a verossimilhança do modelo como

$$D(\beta, \theta, \mathbf{u}) = -2\ln(p(\mathbf{y}|\beta, \theta, \mathbf{u})) = -2 \sum_{i=1}^n \sum_{j=1}^k \ln P(Y_{ij} = y_{ij}|\beta, \theta, \mathbf{u}), \quad (\text{B.1})$$

onde y_{ij} denota a resposta observada (0 ou 1) de Y_{ij} , a variável resposta do avaliado i no item j no modelo TRI-NA ou a variável resposta da observação i no modelo de regressão em que j pode não estar presente o pode corresponder a réplicas das observações. β são os parâmetros associados aos itens no modelo TRI-NA ou os coeficientes de regressão, θ corresponde aos parâmetros associados á função de ligação que estão presentes exclusivamente no modelo de regressão, e \mathbf{u} são os parâmetros associados as habilidades dos avaliados que estão presentes exclusivamente no modelo TRI-NA. O Desvio bayesiano é por sua vez uma medida de bondade de ajuste. Dado dos modelos alternativo a serem comparados, o modelo que melhor ajusta os dados é aquele que tem menor valor do Desvio bayesiano. Mas sua distribuição posterior também pode ser usada para avaliar o ajuste (relativo) de um modelo dado como proposto em vários trabalhos, entre eles Gilks *et al.* (1996) e Spiegelhalter *et al.* (1996). Assim, a quantidade $E[D(\beta, \theta, \mathbf{u})]$, baseada nas distribuições a posteriori de β, θ e \mathbf{u} , é denominada o *esperado do desvio bayesiano a posteriori ou desvio médio a posteriori* (posterior mean of the deviance)

também é uma medida de bondade de ajuste. Esta esperança pode ser aproximada usando o algoritmo MCMC através de

$$E[D(\beta, \theta, \mathbf{u})] \approx \frac{1}{G} \sum_{i=1}^G D(\beta^g, \theta^g, \mathbf{u}^g), \quad (\text{B.2})$$

em que o índice g representa a g -ésima realização simulada de um total de G realizações simuladas. Esta estimativa é denominado como $Dbar$ em Spiegelhalter *et al.* (2002) e é obtido como modo default em WinBugs.

De outra maneira, Carlin e Louis (2000) e Brooks (2002) tem sugerido penalizar $E[D(\beta, \theta, \mathbf{u})]$ como no caso do critério de informação bayesiana (Schwarz) ou o critério de informação de Akaike (Akaike, 1973). Isto leva ao Esperado do Critério de Informação de Akaike ($EAIC$) definido por

$$EAIC = E[D(\beta, \theta, \mathbf{u})] + 2p, \quad (\text{B.3})$$

e ao Esperado do Critério de informação (Schwarz) bayesiana ($EBIC$) definido por

$$EBIC = E[D(\beta, \theta, \mathbf{u})] + p \log N \quad (\text{B.4})$$

no qual p é o número de parâmetros no modelo e N é o número total de observações e \log corresponde ao logaritmo neperiano. Estes critérios podem ser aproximados usando o algoritmo MCMC através de

$$\widehat{EAIC} = Dbar + 2p, \quad \widehat{EBIC} = Dbar + p \log N \quad (\text{B.5})$$

Infelizmente, no caso dos modelos hierárquicos temos problemas com estes critérios, pois não é fácil definir p e N . Por exemplo, na TRI nós temos respostas de k ítems para os indivíduos $i = 1, \dots, n$, então, $N = k \times n$ o total de observações ou $N = n$ o total de indivíduos?. Se as respostas aos ítems em cada indivíduo foram independentes, então a primeira escolha é mais apropriada, mas se as respostas foram perfeitamente correlacionados em cada indivíduo, nós podemos escolher a última. Na TRI nós consideramos o primeiro caso devido a suposição de independência condicional. A situação real é algo intermediário entre estes dois casos. Similarmente, se nós manejamos uma coleção de n efeitos aleatórios como acontece com as variáveis latentes na TRI-NA, uma por cada indivíduo, como isto contribui para o número de parâmetros p ? Se os efeitos aleatórios não tem nada em comum (i.e., eles se comportam essencialmente como os efeitos fixos), eles deveriam contribuir com todo os n parâmetros para p , mas se os dados (ou prioris) indicadam que eles foram todos essencialmente idênticas, eles podem contribuir um pouco mas do que um “número efetivo de parâmetros” para o tamanho do modelo p .

Para direcionar este problema Spiegelhalter *et al.* (2002) propuseram ρ_D como um valor apropriado para penalizar $E[D(\beta, \theta, \mathbf{u})]$. O DIC é uma modificação do esperado do desvio Bayesiano e é definido como

$$DIC = E[D(\beta, \theta, \mathbf{u})] + \rho_D, \quad (\text{B.6})$$

em que

$$\rho_D = E[D(\beta, \theta, \mathbf{u})] - D[E(\beta), E(\theta), E(\mathbf{u})]. \quad (\text{B.7})$$

é denominado como o *número de parâmetros efetivos* (Spiegelhalter *et al*, 1996) e $D[E(\beta), E(\theta), E(\mathbf{u})]$ é o *desvio bayesiano das médias a posteriori ou desvio bayesiano dos esperados a posteriori* (deviance of posterior mean) obtido considerando os valores médios o esperados a posteriori dos parâmetros do modelo. As esperanças acima são baseadas nas distribuições a posteriori de β , θ e \mathbf{u} a qual pode ser aproximada usando o algoritmo MCMC através de

$$D[E(\beta), E(\theta), E(\mathbf{u})] \approx D\left(\frac{1}{G} \sum_{i=1}^G \beta^g, \frac{1}{G} \sum_{i=1}^G \theta^g, \frac{1}{G} \sum_{i=1}^G \mathbf{u}^g\right), \quad (\text{B.8})$$

em que como antes o índice g representa a g -ésima realização simulada de um total de G realizações simuladas. Esta estimativa é denominado como *Dhat* em Spiegelhalter *et al.* (2002) e é obtido como modo default em WinBugs.

O *DIC* penaliza o esperado do desvio bayesiano pela complexidade do modelo ou número de parâmetros efetivos representado por ρ_D . O *DIC* é aproximadamente análogo ao *AIC* e *EAIC* e pode ser aproximado por:

$$\widehat{DIC} = Dbar - Dhat = Dbar + 2\rho_D, \quad (\text{B.9})$$

em que valores baixos do *DIC* indicam o melhor ajuste. Outra vantagem do *DIC* é que como esta baseado na função de verossimilhança, esta pode ser descomposta para cada observação de onde se pode implementar residuais bayesianos (Gelfand, 1996) para avaliar a bondade de ajuste local. A estimativa do *DIC* é obtida como modo default em WinBugs.

Spiegelhalter *et al.* (2002) indicam que o *DIC* como implementado no WinBUGS pode ser usado para comparar modelos complexos (veja-se por exemplo Johnson, 2003) e grandes diferenças nos critérios podem ser atribuídas a diferenças preditivas reais nos modelos. Na modelagem hierárquica a verossimilhança (ou “complexidade do modelo”) não é única no sentido que tem varias versões da verossimilhança que podem ser usadas como acontece nos modelos propostos na teses ao se usar variáveis latentes auxiliares. Portanto o desvio bayesiano do modelo não é único (e também ou *DIC* e ρ_D que são baseados nela) e pode ser calculado de várias modos (ver Delorio e Roberts, 2002). Este fato é importante quando usado variáveis latentes auxiliares.

Levar em conta a complexidade do modelo é particularmente importante na TRI-NA em regressão, a qual introduz um considerável número de variáveis latentes auxiliares \mathbf{Z} . Considerando estas variáveis nós usamos a versões de verossimilhança completa com as variáveis

observadas e as variáveis latentes auxiliares introduzidas (como efeitos fixos e aleatórios na modelagem hierárquico) para obter as distribuições a posteriori para o parâmetros de interesse. Neste caso, *DICs* marginais para as variáveis observadas (efeitos fixos) e variáveis latentes auxiliares (efeitos aleatórios) são apresentadas. Para uma comparação apropriada dos modelos propostos, nos consideramos o *DIC* marginal para as variáveis observadas porque o foco do análise está em $p(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\theta})$ e embora as variáveis latentes auxiliares são introduzidas (em duas etapas) elas não são o foco do análise.

Em tanto tem acordo em relação ao uso do *esperado do desvio bayesiano* (e por tanto de $Dbar$) não se tem acordo em relação ao uso do *desvio bayesiano dos esperados* (veja discussão em Spiegelhalter *et al.*, 2003). Assim Spiegelhalter *et al.*, (2003) advertem que valores de $Dhat$ (e, por consequência, ρ_D e *DIC*) podem ser obtidos dependendo da parametrização usada para a distribuição a priori. Subseqüentes pesquisas no uso apropriado das parametrizações na TRI podem levar a um uso adequado do *DIC*. Também Spiegelhalter *et al.* (2003) indicam que quando se tem distribuições que não são log-côncavas, é teoricamente possível obter valores de ρ_D que são negativos. Quando isto é o caso trata-se das seguintes situações:

- as verossimilhanças não são log-côncavas (e.g. distribuições t-Student)
- quando existe um conflito substancial entre prioris e os dados;
- quando a distribuição posterior para o parâmetro é simétrica e bimodal, e assim a média posterior é um resumo estatístico não muito adequado que dá grandes valores de desvio.

Nos capítulos 2 a 4 é proposto um procedimento de estimação MCMC usando amostragem de Gibbs ao considerar um esquema de dados aumentados (DAGS) introduzindo variáveis latentes auxiliares \mathbf{Z} e incluso \mathbf{v} : Neste caso nós usamos a verossimilhança para dados aumentados das variáveis observadas \mathbf{y} e as variáveis latentes auxiliares (\mathbf{Z}, \mathbf{v}) (efeitos fixos e efeitos aleatórios na modelagem hierárquica, respectivamente), para obter as distribuições posteriores para os parâmetros de interesse. Isto é a verossimilhança $p(\mathbf{y}, \mathbf{Z}, \mathbf{v}|\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u})$. Quando este é o caso, *DICs* marginais tanto para as variáveis observadas (efeito fixo) quanto para as variáveis latentes auxiliares (efeitos aleatórios) são apresentadas na saída no pacote WinBUGS. Isto é, são apresentados diferentes *DICs* marginais. Por considerando este fato uma comparação apropriada de modelos que usam verossimilhança aumentada com outros modelos que não usam a verossimilhança original, nós consideramos o *DIC* marginal para as variáveis observadas dado que o foco da análise está em $p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\beta})$ porque embora as variáveis aleatórias auxiliares sejam consideradas na estimação elas não são o objetivo da análise.

Na presença de variáveis latentes auxiliares, pode ser mas simple o uso de *EAIC* e *EBIC* porque penalizam o *esperado do desvio bayesiano* usando $2p$ e $p \log N$, em que como é usual, p é o número de parâmetros do modelo, e N é o número de observações os quais não mudam ao se usar a formulação hierárquica.

Tomando em conta esta discussão, e especialmente no caso de modelos hierárquicos complexos como os apresentados na teses é muito conveniente usar mais de um critério para a comparação de modelos. Considerando as versões da verossimilhança aumentada para os modelos discutidos na teses nós propomos usar, semelhante a Albert e Chib (1995), os residuais latentes a posteriori $e_{ij}^* = e_{ij} | \beta_j, \theta_i, y_{ij}, z_{ij}, v_{ij}$ para a os modelos TRI-NA e modelos de regressão. Como em Yan *et al.* (2003) nós propomos usar a *soma de quadrados de residuais latentes a posteriori (SQRL)* como uma medida de discrepância global para a comparação dos modelos, em que (SQRL) é definida para o conjunto de dados como

$$SQRL = \sum_{i=1}^n \sum_{j=1}^k e_{ij}^* \quad (\text{B.10})$$

Os valores de individuais de SQRL, isto é, os residuais latentes são interessantes porque podem ser usados como medidas de discrepância de uma observação e sua distribuição posterior pode ser comparada com sua distribuição a priori. Valores específicos destas distribuições são apresentados nos capítulos correspondentes.

Apêndice C

Relações entre os modelos probito assimétricos e probito

Denotamos por \mathcal{M}_1 o modelo probito e \mathcal{M}_2 o modelo probito assimétrico generalizado. As suposições sobre estes modelos essencialmente implicam que os modelos são encaixados, isto é $\mathcal{M}_1 \subset \mathcal{M}_2$. Para ver isto, nós apresentamos nas Proposições a seguir algumas relações entre a verossimilhança dos modelos \mathcal{M}_1 e \mathcal{M}_2 usando versões da verossimilhança baseada no primeiro esquema de dados aumentados introduzidos no Capítulo 3, isto é a *função de verossimilhança de dados aumentados* para o modelo probito assimétrico generalizado dada na expressão (3.14) e a função de verossimilhança de dados aumentados proposta por Albert e Chib (1993) para o modelo probito dada na expressão (2.8).

Usando o algoritmo de dados aumentado como considerado em Albert e Chib (1993) para o modelo probito, em que variáveis latentes subjacentes $\mathbf{Z} = (Z_1, \dots, Z_n)$ são $Z_i \sim N(\mathbf{x}'_i\boldsymbol{\beta}, 1)$, e seja $D = (n, \mathbf{y}, \mathbf{X}, \mathbf{z})$ que denota os “dados aumentados”, temos que, a verossimilhança de dados aumentados de \mathcal{M}_1 pode ser escrita como

$$L(\mathcal{M}_1) = L(\boldsymbol{\beta}|D) = \prod_{i=1}^n \phi(z_i - \mathbf{x}'_i\boldsymbol{\beta}) I(z_i, y_i) \quad (\text{C.1})$$

onde, como antes, $\phi(\cdot)$ e $\Phi(\cdot)$ denotam a fdp e fda da normal padrão respectivamente e $I(\cdot)$ é a função indicadora definida antes.

Também, substituindo a fda da normal assimétrica em (3.14), nós obtemos que

$$L(\mathcal{M}_2) = 2^n \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}'_i\boldsymbol{\beta} + \mu}{\sigma}\right) \Phi\left(-\lambda\left(\frac{z_i - \mathbf{x}'_i\boldsymbol{\beta} + \mu}{\sigma}\right)\right) I(z_i, y_i) \quad (\text{C.2})$$

Por facilidade nós escolhemos o *BBB probito assimétrico* definido na proposição 3. Assim, as proposições seguintes indicam interessantes relações entre os modelos *BBB probito assimétrico* e probito.

Proposição C.1. A relação entre a verossimilhança aumentada dos modelos *BBB probito as-*

simétrico e probito é dada por

$$\log(L(\mathcal{M}_2)|L(\mathcal{M}_1)) = n\log 2 + \sum_{i=1}^n \log \Phi(-\lambda(z_i - \mathbf{x}'_i \boldsymbol{\beta})) \quad (\text{C.3})$$

Prova. Considere, como na Proposição 3, $\theta = (0, 1, \lambda)$ em (C.2). Note que $I(z_i, y_i)$ em (C.1) é diferente de $I(z_i, y_i)$ em (C.2) porque z_i em (C.2) é distribuído como a normal-padrão e z_i em (C.1) é normalmente distribuída. Mas como $\Phi(\lambda z_i)$ é positivo para qualquer z_i , segue que $I(z_i, y_i)$ em (C.1) e (C.2) são iguais quando avaliados no mesmo $z_i \sim N(\mathbf{x}'_i \boldsymbol{\beta}, 1)$ e y_i . Por tanto, nós temos que

$$L((\boldsymbol{\beta}, \lambda)|\mathbf{Z}, \mathbf{y}) = 2^n \prod_{i=1}^n \Phi(-\lambda(z_i - \mathbf{x}'_i \boldsymbol{\beta})) L(\boldsymbol{\beta}|\mathbf{Z}, \mathbf{y}) \quad (\text{C.4})$$

e o resultado segue tomando logaritmo nesta expressão. \square

Existem vários modos de medir a distância entre modelos. Nós consideramos a divergência de Kullback-Leibler que não é propriamente uma distância mas é não negativa e alias, não simétrica.

Proposição C.2. A divergência de Kullback-Leibler entre os modelos probito e BBB pa é dada por

$$KL_{1,2} = -n\log 2 - \sum_{i=1}^n \int \log \Phi(-\lambda t_i) \phi(t_i) dt_i \quad (\text{C.5})$$

em que $t_i = z_i - \mathbf{x}'_i \boldsymbol{\beta}$.

Prova. Usando (C.3) e usando a definição da divergência de Kullback-Leibler nós temos que

$$KL_{1,2} = E_1 \left(\log \frac{f(\mathbf{y}, \mathbf{z} | (\boldsymbol{\beta}))}{f(\mathbf{y}, \mathbf{z} | \boldsymbol{\beta}, \lambda)} \right) = -n\log 2 - \sum_{i=1}^n E_1 [\log \Phi_i(-\lambda(z_i - \mathbf{x}'_i \boldsymbol{\beta}))],$$

em que

$$E_1 [\log \Phi(-\lambda(z_i - \mathbf{x}'_i \boldsymbol{\beta}))] = \int \log \Phi(-\lambda(z_i - \mathbf{x}'_i \boldsymbol{\beta})) \phi(z_i - \mathbf{x}'_i \boldsymbol{\beta}) dz_i = \int \log \Phi(-\lambda t_i) \phi(t_i) dt_i \square$$

Note que a medida de divergência KL não depende dos valores observados x_i e dos parâmetros $\boldsymbol{\beta}$. A Figura 1 mostra a expressão $KL_{1,2}$ para diferentes valores de λ e o número de observações n usando integração por Monte Carlo. Observe que a divergência entre os modelos probito e BBB pa cresce com $|\lambda| > 0$ e n .

Observação 1. Também $KL_{1,2}$ pode ser escrita como $KL_{1,2} = \sum_{i=1}^n \int Z(t_i, \lambda) \phi(t_i) dt_i$, em que $Z(t_i, \lambda) = \log(2\Phi(-\lambda t_i))$ é a função definida na libreria *sn* no programa *R*, e pode ser estimada

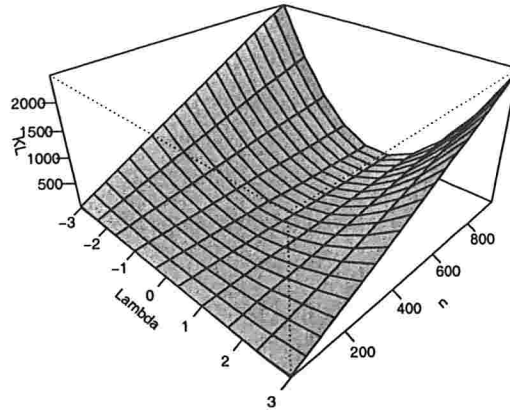


Figura C.1 Divergência de Kullback-Leibler entre os modelos probito e BBB pa para diferentes valores de λ e n

por $\widehat{KL}_{1,2} = \sum_{j=1}^m \sum_{i=1}^n \frac{Z(t_i, \lambda)}{m}$ com $t_i \in N(0, 1)$ e $m \rightarrow \infty$.

Outro modo de comparar \mathcal{M}_1 e \mathcal{M}_2 é fazendo testes de hipóteses de \mathcal{M}_1 em relação a \mathcal{M}_2 o qual pode ser feito usando o Factor de Bayes (Kass e Raftery, 1995). Suponha que nós estamos comparando os modelos \mathcal{M}_1 e \mathcal{M}_2 , então, o modelo \mathcal{M}_1 deve ser mais em acordo com os dados D , se $\pi(\mathcal{M}_1|D) > \pi(\mathcal{M}_2|D)$. Conseqüentemente, considerando as duas hipóteses H_1 e H_2 que correspondem a suposição de dois modelos alternativos, o probito \mathcal{M}_1 e o modelo BBB pa \mathcal{M}_2 para os dados D , temos que o factor de Bayes em favor de H_1 (e contra H_2) é dado pela distribuição posterior para a razão a priori

$$BF_{12} = \frac{p(D|\mathcal{M}_1)}{p(D|\mathcal{M}_2)} = \frac{\int L(\mathcal{M}_1)\pi(\beta|\mathcal{M}_1)d\beta}{\int L(\mathcal{M}_2)\phi(\beta, \lambda|\mathcal{M}_2)d\beta d\lambda}$$

Considerando esta definição, a seguinte proposição é uma conseqüência direta.

Proposição C.3. O Factor de Bayes para testar a hipóteses em favor do modelo probito \mathcal{M}_1 quando é considerado o modelo BBB pa é dado por

$$BF_{12} = \prod_{i=1}^n \frac{\int \phi(z_i - x_i'\beta)I(z_i, y_i)\pi(\beta)d\beta}{\int f_{\theta}(z_i)I(z_i, y_i)\pi(\beta)\pi(\lambda)d\beta d\lambda}$$

onde $\theta = (x_i'\beta, 1, -\lambda)$. Em particular, se β é conhecido, o Factor de Bayes se reduz a seguinte

expressão

$$BF_{12}^* = \frac{1}{\prod_{i=1}^n \int 2\Phi(-\lambda(z_i - \mathbf{x}'_i\boldsymbol{\beta}))\pi(\lambda)d\lambda}.$$

A expressão BF_{12} não tem forma fechada e cálculos numéricos podem ser complexos. Mas, quando $\boldsymbol{\beta}$ é conhecido, BF_{12}^* tem uma forma simples. Os seguintes resultados são conseqüências direta da última expressão;

- Se $\lambda \sim N(0, 1)$ então $2\Phi(-\lambda(z_i - \mathbf{x}'_i\boldsymbol{\beta}))\phi(\lambda) = f_\theta(\lambda)$, em que $\theta = (0, 1, z_i - \mathbf{x}'_i\boldsymbol{\beta})$ corresponde à distribuição normal assimétrica como uma função de λ . Segue que $BF_{12} = 1$ e assim o Factor de Bayes não pode distinguir entre os modelos probito e *BBB pa*.
- Em general, se $\lambda \sim N(m, v^2)$ então

$$BF_{12}^* = \frac{1}{2^n \Phi_n[-m(z_i - \mathbf{x}'_i\boldsymbol{\beta}), I_n + v^2(z_i - \mathbf{x}'_i\boldsymbol{\beta})(z_i - \mathbf{x}'_i\boldsymbol{\beta})']}$$

em que $\Phi_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é a fda da normal multivariada $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note que se $m = 0$ o resultado particular acima segue. Este resultado é semelhante ao resultado encontrado por Vidal et al. (2004) para comparar os modelos normal e normal assimétrico.

- Se $\lambda^2 \sim G(a, b)$ and $\lambda > 0$, então

$$BF_{12}^* = \frac{1}{2^n T_n[(z_i - \mathbf{x}'_i\boldsymbol{\beta})\sqrt{\frac{a}{b}}, I_n, 2a]}$$

onde $T_n(\cdot, \boldsymbol{\Sigma}, \nu)$ é a fda da $t_n(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$ multivariada e I_n e a matriz identidade de ordem n . Também este resultado é similar ao resultado encontrado por Vidal et al. (2004) e se um sinal para λ pode ser assumido nós podemos considerar uma priori dada a informação fornecida pelos dados. Expressões similares são obtidas quando $\lambda < 0$ e $\lambda^2 \sim G(a, b)$ é assumido.

A importância de estes resultados es que o uso de prioris normais para λ não permite distinguir o modelo probito do modelo *BBB pa*.

Apêndice D

Programas para a implementação dos modelos propostos desenvolvidos no WinBUGS

D.1 Estudo de eficiência nos Modelos PN e PAN

Este programa corresponde aos modelos PN e PAN nas suas versões DAGS considerando o segundo método de dados aumentados no modelo PAN e as parametrizações d e λ para os dados de Tanner (1996).

```
model{
  for (i in 1:n) {
    for (j in 1:I) {
      m[i,j] <- a[j]*u[i] - b[j]

      #PN model
      # z[i,j]~dnorm(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1])
      #   resid[i,j]<-z[i,j]-m[i,j]
      #SPN model
      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]-m[i,j]
    }
  }
  #residuals
  res2[i,j]<-pow(resid[i,j],2)
}

#priors for item parameters
for (j in 1:I) {
  b[j] ~ dnorm(0,0.5)
  bc[j] <- b[j] - mean(b[])
  a[j] ~ dnorm(1,2)I(0,)
  ac[j]<-a[j]/pow(a[1]*a[2]*a[3]*a[4]*a[5],0.2)
  beta[j]<-bc[j]/a[j]
}

#SPN, SPSN model
```

```

# delta parametrization
      d[j]~ dunif(-1,1)
      preczs[j]<- 1/(1-pow(d[j],2))
      lambda[j]<-d[j]*sqrt(preczs[j])
#lambda parametrization
#      lambda[j] ~ dt(0,2,2)
#      preczs[j]<- 1+pow(lambda[j],2)
#      d[j]<- lambda[j]/sqrt(preczs[j])
}

#priors for latent variable in PN model
for(i in 1:n){ u[i]~ dnorm(0,1)}

#PN model #
lo[1]<- -50; lo[2]<- 0; ## i.e., Z*y=0~N(m,1)I(-50,0)
#up[1]<- 0; up[2]<-50; ## i.e., Z*y=1~N(m,1)I(0,50) \\
#SPN, SPSN model#\
lo[1]<- -50; lo[2]<- 0; ## i.e., Z*y=0~N(-d*v+m,1-d^2)I(-50,0)
up[1]<- 0; up[2]<-50; ## i.e., Z*y=1~N(-d*v+m,1-d^2)I(0,50)

for(j in 1:I) { resmean[j]<-sum(res2[,j])} sse<-sum(resmean[])
mu<-mean(u[]) du<-sd(u[])
}

Dados: TANNER
Inits list(b = c(0,0,0,0,0,0), a =c(1,1,1,1,1,1))\\
Inits list(b =c(0,0,0,0,0,0), a =c(1,1,1,1,1,1), d=c(0,0,0,0,0,0))
\\Inits list(b = c(0,0,0,0,0,0), a =c(1,1,1,1,1,1),
lambda=c(0,0,0,0,0,0))

Data list( n=39, I=6,
          y= structure(.Data =
            c(1,0,1,0,0,1,
              1,1,1,0,1,1,
              1,0,0,0,1,0,
              0,1,0,0,0,0,
              1,1,1,0,0,1,
              0,0,1,0,1,1,
              1,1,1,1,0,0,

```

```
1,1,1,1,1,1,
0,1,0,0,1,1,
0,1,1,0,1,1,
0,1,0,0,0,0,
0,1,0,1,0,1,
0,1,0,0,0,1,
1,0,1,0,0,1,
1,0,1,0,0,0,
0,1,0,0,0,0,
0,1,1,0,1,1,
0,0,1,1,0,1,
1,1,1,0,1,1,
0,0,1,0,0,0,
1,1,0,0,0,1,
0,1,0,0,0,0,
0,1,1,1,0,0,
1,1,1,1,1,1,
1,0,0,0,0,0,
1,0,0,0,0,0,
0,1,0,0,0,1,
1,1,1,1,1,0,
1,0,0,1,1,0,
1,1,0,0,0,0,
1,0,0,0,0,0,
0,0,1,1,0,0,
0,0,0,1,1,0,
0,0,0,0,0,1,
1,0,0,0,0,0,
0,1,0,0,1,1,
1,0,0,1,1,1,
1,1,1,1,1,1,
1,0,0,0,1,1),
.Dim = c(39,6))
```

)

D.2 Análise de sensibilidade na especificação de prioris no modelo PAN

Este programa corresponde ao estudo de sensibilidade da escolha de prioris para o modelo PAN usando o segundo método de aumentação e a parametrização d para os dados de Matemática.

```
model{
  for (i in 1:n) {
    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)
    }
  }

  #priors for item parameters
  for (j in 1:I) {
    #prior I: Sahu (2002)
    a[j] ~ dnorm(1,2)I(0,)
    b[j] ~ dnorm(0,0.5)
    d[j] ~ dunif(-1,1)

    #prior II: Sahu (2002)
    #a[j] ~ dnorm(1,2)I(0,)
    #b[j] ~ dnorm(0,0.5)
    #d[j] ~ dnorm(0,1)I(-1,1)

    #prior III: Patz e Junker (1999)
    # a[j] ~ dlnorm(0,2)
    # b[j] ~ dnorm(0,0.5)
    # d[j] ~ dunif(-1,1)

    #prior IV Patz e Junker (1999)
    # a[j] ~ dlnorm(0,2)
    # b[j] ~ dnorm(0,0.5)
    # d[j] ~ dnorm(0,1)I(-1,1)

    preczs[j]<- 1/(1-pow(d[j],2))
    lambda[j]<-d[j]*sqrt(preczs[j])
  }
}
```

```

    }
#priors latent variable
  for(i in 1:n){ u[i]~ dnorm(0,1)  }

lo[1]<- -50; lo[2]<- 0; ## i.e.,
Zs*|y=0~N(-delta*V+m,1-delta^2)I(-50,0) up[1]<- 0; up[2]<-50;
## i.e., Zs*|y=1~N(-delta*V+m,1-delta^2)I(0,50)

  for(j in 1:I) {    resmean[j]<-sum(res2[,j])}
    sse<-sum(resmean[])
    mu<-mean(u[])
    du<-sd(u[])
  }

Dados Matem\ '{a}tica

Inits list(b = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0), a
=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1), d=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0))

Data list(n=131, I=14,
          y= structure(
            .Data =  c(
1,1,0,1,1,0,1,1,1,1,1,0,0,1, 1,1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,1,0,1,0,0,1,0,1,0,0,0,1, 1,0,0,1,1,0,1,0,1,1,0,0,1,0,
1,1,1,1,1,0,1,1,1,1,1,1,1,1, 1,0,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,1, 1,1,0,1,0,0,1,1,1,1,1,1,1,1,
0,1,1,1,1,1,1,0,1,1,1,1,1, 1,0,1,1,1,0,1,0,1,0,1,0,1,1,
1,0,0,1,1,0,0,1,1,1,1,0,1,1, 1,1,0,1,0,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1, 1,0,0,1,1,1,0,1,0,1,0,0,1,1,
1,1,1,1,1,1,1,1,1,0,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,
0,1,1,1,1,0,1,0,1,1,1,1,0,1, 1,1,0,1,1,0,1,1,1,0,1,0,1,1,
1,0,0,1,1,0,1,1,1,1,1,0,0,1, 1,1,0,1,1,0,1,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,0,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,0,0,
1,0,0,1,1,0,1,1,1,1,1,0,1,1, 0,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1, 1,0,0,1,0,0,1,1,1,1,1,1,1,1,
1,1,1,0,1,1,1,1,1,1,1,1, 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, 1,1,0,1,1,0,1,1,1,0,1,1,1,1,
1,1,1,1,1,0,1,1,1,1,1,1,1,1, 1,0,1,1,1,0,1,1,1,1,1,1,0,1,
1,1,1,1,1,1,1,1,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,0,1,1,0,1,1,
1,1,1,1,1,1,1,1,0,1,1,0,1,1, 1,1,0,1,1,0,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1, 1,1,1,1,1,1,1,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,0,1,1,0,1,1, 1,1,0,1,1,1,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1, 0,1,1,1,0,0,1,1,1,1,0,0,1,1,
0,1,0,1,1,0,1,1,1,0,1,0,0,1, 1,1,0,1,1,1,1,1,1,0,1,0,1,1,
1,1,1,1,1,0,1,1,0,1,1,1,1,1, 1,1,1,1,1,0,1,1,1,1,1,0,0,1,
1,1,0,1,1,1,1,1,1,1,1,1,1,1, 0,1,1,1,0,0,1,1,1,1,1,1,1,1,
1,1,0,1,1,0,0,1,1,1,1,0,1,1, 1,1,1,1,1,1,1,1,1,1,1,0,1,1,
0,0,0,1,1,1,1,1,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,1,1,1,1,1,1,
1,1,0,1,0,0,1,0,1,1,1,0,0,1, 1,1,1,1,1,0,1,1,1,1,1,1,0,1,
0,1,0,1,1,0,1,1,1,1,1,0,0,1, 0,1,0,1,1,0,0,0,0,0,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,1,1,1,1,0,1, 1,1,1,1,1,0,1,1,0,1,1,1,1,
0,1,1,1,1,1,1,1,1,1,1,0,0,1, 1,1,1,1,1,0,0,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,0,1,1,1,1, 1,1,1,1,1,0,1,1,1,1,1,0,0,
1,1,1,1,1,0,1,1,0,1,1,0,1,1, 1,1,0,1,1,0,1,1,0,1,1,0,1,1,
0,1,0,0,0,0,1,1,1,1,1,0,0,1, 1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,1,1,1,0,1,1,1,1,1,0,0,1, 1,1,1,1,1,0,1,0,1,1,1,0,1,1,
0,1,1,1,1,0,1,0,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,1,1,0,1,1,
0,1,0,1,0,1,1,1,1,1,0,1,1,1, 1,1,1,1,1,1,1,1,1,0,1,1,1,
0,1,1,1,1,1,1,1,1,1,1,0,1,0, 1,1,1,1,1,1,0,1,1,1,1,0,1,1,
1,1,0,1,1,0,1,1,1,1,1,0,1,1, 1,1,1,1,1,1,1,1,1,0,1,1,1,
1,1,0,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,0,1,1,
1,0,1,1,1,0,1,1,0,0,1,0,1,1, 1,1,1,1,0,1,1,1,0,1,1,0,1,1,
1,1,1,1,1,1,1,1,0,1,1,1,1,1, 1,1,0,1,1,0,0,1,1,0,0,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1, 0,1,0,0,1,1,1,0,1,1,1,1,1,1,
1,1,0,1,1,0,1,1,1,1,1,1,1, 1,1,0,1,1,1,1,1,1,1,1,0,1,1,
1,0,1,1,1,0,1,1,0,1,1,0,0,1, 0,1,1,1,1,0,1,1,1,1,1,0,1,1,
1,1,0,0,0,0,1,1,1,1,0,0,0,1, 1,1,0,0,1,0,1,0,1,0,0,0,0,0,
1,1,0,1,1,0,1,1,1,1,1,0,1,1, 1,1,0,1,1,1,1,1,1,1,0,1,1,
0,1,0,1,1,0,1,1,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,1,1,1,1,1,
0,1,0,1,1,1,1,1,1,1,1,0,1,1, 0,1,1,1,1,0,1,1,1,0,1,0,1,1,
1,1,0,1,0,1,1,1,1,1,1,0,1,1, 1,0,0,0,1,1,1,1,0,1,1,0,1,0,
1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,0,1,1,1,1,1,1,1,
1,1,1,1,0,0,1,1,0,1,1,0,0,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,

```
1,1,0,1,1,0,1,1,1,1,1,0,1,1, 1,1,0,1,1,1,1,0,1,1,0,1,1,  
0,0,1,1,1,0,1,1,1,1,1,1,1,1, 1,1,0,1,1,0,1,1,1,1,1,0,1,1,  
1,1,1,1,1,1,1,1,0,1,1,0,1,1, 0,1,0,1,0,1,0,0,1,1,1,0,0,1,  
0,1,1,0,1,0,1,1,1,1,1,0,1,1, 1,1,0,1,1,1,1,1,1,1,1,1,0,  
1,1,0,1,1,0,1,1,1,1,1,0,1,1, 1,1,1,1,1,0,1,1,1,1,1,0,1,  
0,1,0,1,1,0,1,0,1,0,1,0,1,1, 0,0,0,1,0,0,1,0,1,0,1,1,1,1,  
1,1,1,1,1,0,1,1,0,1,1,0,1,1, 1,1,0,1,1,0,1,1,1,1,1,0,1,1,  
0,1,1,1,1,0,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,0,0,0,1,0,0,1,  
1,0,0,1,1,0,1,1,1,0,1,1,1,1, 1,1,0,1,1,0,1,1,1,1,1,0,1,1,  
0,1,0,0,1,0,0,0,1,1,1,0,1,1, 1,1,0,1,1,1,1,1,1,1,0,1,1,  
1,1,0,1,1,0,1,0,0,1,1,0,1,1, 1,0,0,1,1,1,1,1,0,1,1,0,1,1,  
1,1,0,1,1,1,1,1,1,0,1,0,1,1, 1,1,0,1,1,0,1,1,0,1,1,0,1,1,  
1,0,0,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,0,1,1,1,1,1,0,1,1,  
1,1,0,1,0,0,1,1,0,0,1,0,1,1),  
    .Dim = c(131,14))  
))
```

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. e Chib, S. (1993). Bayesian Analysis of binary and polytomous response data. *Journal of the American Statistical Association*, 88, 669-679.
- [3] Albert, J.H. e Ghosh, M. (2000). Item response modeling. *Generalized Linear Models: A Bayesian Perspective*(D. Dey, S. Ghosh & Mallick, eds.), Marcel-Dekker, New York, 173-193.
- [4] Andrade, D. F. , Tavares, H. R. e Valle, R. C. (2000). Introdução à Teoria da resposta ao Item : Conceitos e Aplicações. 14º SINAPE : Caxambu, MG.
- [5] Akaike, H. (1973). Information Theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csáki (Eds.) 2nd International Symposium on Information Theory (pp. 267-281). Budapest: Akadémiai Kiadó.
- [6] Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, 68, 357-363.
- [7] Arellano-Valle, R., Bolfarine, H., e Iglesias, P. (1994). Preditivistic Interpretation to the multivariate t distribution. *Test*, 3, 1-16.
- [8] Arnold, B.C., Beaver, R.J., Groeneveld, R., e Meeker, W. Q. (1993). . *Psychometrika* 58, 471-488
- [9] Arnold, B. C. e Beaver, R. J., (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Sociedad de Estadística e Investigación Operativa Test* 11, 7-54.
- [10] Assunção, C. (1999). Estimación dos parâmetros de modelos da teoria de resposta ao item e aplicações. Mestrado em Estatística - UFMG.
- [11] Azzalini A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statistical*, 12, 171-178.

- [12] Azzalini A. e Capitanio A. (1999). Statistical applications of the multivariate skew normal distributions. *J. R. Statist. Soc. B*, 61, 579-602.
- [13] Azzalini A. e Dalla Valle A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715-726.
- [14] Baker, F.B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [15] Bartholomew, D.J. e Knoot, M. (1999). *Latent variable models and factor analysis*. (2nd ed.). London: Arnold. (Kendall's Library of Statistics 7).
- [16] Basu, S. e Mukhopadhyay, S. (2000). Binary Response Regression With Normal Scale Mixtures Links, in *Generalized Linear Models: A Bayesian Perspective*, eds. D.K. Dey, S.K. Ghosh, and B.K. Mallick, New York: Marcel Dekker.
- [17] Bazán, J., Espinosa G. e Farro Ch. (2002). Rendimiento y actitudes hacia la matemática en el sistema escolar peruano. In Rodriguez, J., Vargas, S., (eds.) *Análisis de los Resultados y Metodología de las pruebas Crecer 1998. Documento de trabajo 13*. Lima: MECEP-Ministerio de Educación. 55-70.
- [18] Bazán, J., Branco, M. e Bolfarine, H. (2004a). A skew item response model. ISBA 2004 World Meeting. Viña del Mar. Chile, May 23-27, 2004. ISBA (International Society for Bayesian Analysis).
- [19] Bazán, J., Bolfarine, H., e Branco, D. M. (2004b). A new family of asymmetric models for item response theory: A SKEW-NORMAL IRT FAMILY. 49p. *Technical report (RT-MAE-2004-17)*. Department of Statistics. University of São Paulo.
- [20] Bazán, J., Bolfarine, H. e Aparecida, R. (2004a). Bayesian estimation via MCMC for probit-normal model in item response theory. 26p. *Technical report (RT-MAE-2004-15)*. Department of Statistics. University of São Paulo.
- [21] Bedrick, E. J., Christensen, R. e Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91, 1450-1460.
- [22] Béguin, A. A. e Glas, C. A. W. (2001). MCMC estimation of multidimensional IRT models. *Psychometrika*, 66, 541-562.
- [23]) Best N. G., Cowles M. K. e Vines, S, K, (1997) *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.4*. [Computer program]. MRC Biostatistics Unit, Cambridge University.

- [24] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord & M. R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA : Addison-Wesley.
- [25] Bliss, C. I. (1935). The calculation of the dose-mortality curve. *Annals Applied Biology*, 22, 134-167.
- [26] Bock, R. D. e Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* 64,153-168.
- [27] Boomsma, A., van Duijn, M. A. J. e Snijders, T. A. B. (2001). *Essays on Item Response Theory*. Springer - Verlag New York, Inc.
- [28] Borsboom, D., Mellenbergh, G. J. e Van Heerden, J.(2003). The Theoretical Status of Latent variables. *Psychological Review*, 110, 203-219.
- [29] Branco, M.D. and Dey, D.K.,(2002). Regression model under skew elliptical error distribution. *The Journal of Mathematical Sciences*, Delhi, New Series, 1, 151-169.
- [30] Branco,M.D. and Dey, D.K. (2001). A class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79, 99-113.
- [31] Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002). *Journal of the Royal Statistical Society Series B*, 64, 3,616-618.
- [32] Carlin, B.P. e Louis, T.A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis Essays on Item Response Theory*. Second edition. New York: Chapman & Hall.
- [33] Chen, M-H, Shao, Q. M, e Ibrahim, J. G (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer Verlag.
- [34] Chen, M-H, Dey, D. K., e Shao, Q-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94, 448, 1172-1186.
- [35] Chen, M-H,,Dey, D.K., e Shao, Q-M. (2001). Bayesian analysis of binary data using skewed logit models. *Calcutta Statistical Association Bulletin*, 51, 201-202.
- [36] Chen, M-H. (2004). The Skewed Link Models for Categorical Response Data, in *Skew-elliptical distributions and their applications: a journey beyond normality*,ed. Marc. G. Genton, Boca Raton: Chapman & Hall/CRC.
- [37] Chen, M-H, e Dey, D. K. (1998). Bayesian Modelling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions. *Sankhyā*, Ser. A. 60, 322-343.

- [38] Chen, M-H e Shao, Q. M. (2000). Property of posterior distribution for dichotomous quantal response model. *Proceedings of the American Mathematical Society*. 129,1,293-302.
- [39] Chib, S. e Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49, 327-335.
- [40] Chib, S. e Greenberg, E. (1995). Bayesian Analysis of Multivariate Probit Models. *Biometrika*, 49, 327-335.
- [41] Czado, C., e Santner, T. J. (1992a). Orthogonalizing link transformation families in binary regression analysis. *Canadian Journal Statistical*, 20, 51-62.
- [42] Czado, C., e Santner, T. J. (1992b). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, 33, 213-231.
- [43] Czado, C. (1994). Bayesian Inference of Binary Regression Models With Parametric Link. *Journal of Statistical Planning and Inference*, 41, 121-140.
- [44] Czado, C. (1997). On selecting parametric link transformation families in generalized linear models. *Journal of Statistical Planning and Inference*, 61, 125-139.
- [45] Czado, C., e Munk, A. (2000). Noncanonical links in generalized lineal models-when is the effort justified?. *Journal of Statistical Planning and Inference*, 87, 317-345.
- [46] Czado, C. e Raftery, A.E. (2001). Choosing the Link Function and Accounting for Link Uncertainty in Generalized Linear Models using Bayes Factors. . Working Paper no 18. Center for Statistics and the Social Sciences University of Washington. Disponível na web.
- [47] Congdon P. (2001). *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, UK.
- [48] Collet, D. (2003). *Modelling Binary Data*. Chapman & Hall/CRC, Second Edition, Boca Raton, USA.
- [49] Dalla Valle, A. (2004). The skew-normal distribution, in *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M. G., Ed., Chapman & Hall / CRC, Boca Raton, FL, pp. 3-24.
- [50] Delorio, M. e Robert, C. P. (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin and van der linde (2002). *Journal Royal Statistics Society. Serie B*. 64, 629-630.
- [51] Dempster, A. P. (1997). The direct use of likelihood for significance testing. *Statistics and Computing* B. 7, 257-252.
- [52] Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society*, 57, 45-97.

- [53] Du Toit, M. (2003). *IRT from SSI*. Scientific Software International.
- [54] Fischer, G. e Molenaar, I. (1995). *Rasch Models. Foundations, recent development, and applications*. The Netherlands: Springer-Verlag.
- [55] Fox, J. P. e Glas, C. A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 271-288.
- [56] Fox, J. P. e Glas, C. A.W. (2003). Bayesian Modeling of measurement error in predictor variables using item response theory. *Psychometrika*, 68, 169-191.
- [57] Froelich, A. G. (2000). *Assesing unidimensionality of test item and some asymptotics of parametric item response theory*. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, Departament of Statistics.
- [58] Gamerman, D. (1997). *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*. Londres : Chapman & Hall.
- [59] Gelfand A. E. (1996). Model determination using sampling-based methods. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.). *Markov Chain Monte Carlo in practice* (pp. 145-161). London., Chapman & Hall.
- [60] Gelman, A., Meng, X.-L. e Stern, H. (1996). Posterior predictive assesment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733-807.
- [61] Gelman, A. e Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
- [62] Genton, M. G. (2004), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Edited Volume, Chapman & Hall CRC, Boca Raton, FL.
- [63] Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1, 141-150.
- [64] Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, 25, 400-405.
- [65] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.). *Bayesian Statistics*, 4, (pp 169-193). Oxford: Oxford University Press.
- [66] Ghosh, M., Ghosh, A., & Chen, Ming-Hui & Agresti, A. (2000). Noninformative priors for one parameter item response models. *Journal of Statistical Planning and Inference*. 88, 99-115.

- [67] Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (1996). Markov Chain Monte Carlo in practice. London : Chapman & Hall.
- [68] Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- [69] Glas, C. A. W. e Meijer, R. (2003). A Bayesian Approach to Person Fit Analysis in Item Response Theory Models. *Applied Psychological Measurement*, 27,217-233.
- [70] Greene, W. (2003). *Econometric Analysis*. 5th Edition. Prentice-Hall.
- [71] Guerrero, V. M., e Johnson, R. A. (1982). Use of the Box-Cox Transformation With Binary Response Models. *Biometrika*, 69, 309-314.
- [72] Hambleton, R.K., Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publishers.
- [73] Hashimoto, Y. (2002). Motivation and willingness to communicate as predictors of reported 12 use: the japanesse context. *Second Language Studies*, 20 (2), 29-70.
- [74] Haro-López, R.A., Mallick, B. K. e Smith, A.F.M. (2000). Binary Regression Using Data Adaptative Robust Link Functions, in *Generalized Linear Models: A Bayesian Perspective*, eds. D.K. Dey, S.K. Ghosh, and B.K. Mallick, New York: Marcel Dekker.
- [75] Henze, N. (1986). A probabilistic representation of the "skew-normal" distribution. *Scandinavian Journal Statistical*. 13, 271-275.
- [76] Holland, P. e Rosenbaum, P. (1986). Conditional Association and Unidimensionality in Monotone Latent variable models. *The Annals of Statistics*. 14, 1523-1543.
- [77] Ibrahim, J. G. e Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffrey's prior. *Journal of the American Statistical Association*. 93, 1282-1293.
- [78] Jackman, S. (2001). Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking. *Political Analysis*, 9 , 227-242.
- [79] Jackman, S. (2004). Bayesian analysis for political research. *Annual Review Political Science*, 7, 483-505.
- [80] Johnson, V. e Albert, J. (1999). *Ordinal Data Modeling*. New York, MA: Springer-Verlag.
- [81] Johnson (2003). On the use of Heterogenous thresholds ordinal regression models to account for individual diferences in response style. *Psychometrika*, 68, 563-583.
- [82] Junker, B. W. e Ellis, J. L. (1997). A characterization of monotone unidimensional latent variable models. *Annals of Statistics*, 25, 1327-1343.

- [83] Kass, R. E. e Raftery, A. E.(1995). Bayes Factor. *Journal of the American Statistical Association*, 93, 1282-1293.
- [84] Kass, R. E., Carlin, B. P., Gelman, A. e Neal, R. (1998). Markov Chain Monte Carlo in practice: A roundtable discussion. *American Statistician*, 52, 93-100.
- [85] Kim, S., Cohen, A. S., Baker, F. B., Subkoviak, M. J. e Leonard, T. (1994). An investigation of hierarchical Bayes procedures in item response theory. *Psychometrika*, 59, 405-421.
- [86] Lachos, V.H. (2004). Modelos Lineares Mistos Assimétricos. Tese de Doutorado. IME-USP.
- [87] Liseo e Loperfido (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statistics Probability Letters*. 395-401.
- [88] Linardakis, M. e Dellaportas, P. (2002). An approach to multidimensional item response modeling. E. I. George. (Ed.). *Bayesian Methods with applications to science policy and oficial statistics*. 331-340.
- [89] Lord, F. M. (1952). *A theory of test scores*. New York: Psychometric Society.
- [90] Lord, F., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Adisson-Wesley.
- [91] Martin, A.D., e Quinn, K. M. (2002). MCMC pack. scythe.wustl.edu/mcmcpack.html.
- [92] Matos, G. S. (2001). *Teoria da Resposta ao Item : Uma Proposta de Modelo Multivariado*. Universidade Federal de Pernambuco - UFPE - Departamento de Estatística.
- [93] Micceri, T. (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures. *Psychological Bulletin*, 105(1), 156-166
- [94] Mislevy, R., J. e Bock, R.D. (1990). *BILOG: Maximum likelihood item analysis and test scoring with logistic models for binary items* [Computer program]. Chicago: Scientific Software International.
- [95] Mislevy, J. M. (1984). Estimating Latent Distributions. *Psychometrika*, Vol. 49, 359 - 381.
- [96] Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- [97] Mislevy, R.J. (1991). Randomization-based inference about latent vaiables from complex samples. *Psychometrika*, 56, 177-196.
- [98] Mislevy, R. J. e Bock, R. D. (1990). *BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Models*. Chicago : Scientific Software, Inc.

- [99] Neal, Radford. (2003). Slice Sampling. *The Annals of Statistics*, 31(3), 705-767.
- [100] Neal R. (1998). Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In *Learning in Graphical Models*, (M I Jordan, ed). Kluwer Academic Publishers, Dordrecht, pp. 205-230. <http://www.cs.utoronto.ca/radford/publications.html>
- [101] Nojosa, R. T. (2001). *Modelos Multidimensionais na Teoria da Resposta ao Item*. Dissertação de Mestrado. Universidade Federal de Pernambuco - UFPE - Departamento de Estatística.
- [102] Ozán, N.S. (2003). Modelo assimétrico com erros nas variáveis. tese de Doutorado. IME.
- [103] Paulino, C., Soares, P. e Neuhaus, J. (2003). Binomial Regression with Misclassification. *Biometrics*, 59,670-675.
- [104] Parrish, R. e Bargmann, R. (1981) A method for evaluation of cumulative probabilities of bivariate distributions using the pearson family. *Statistical Distributions in Scientific Work*, 5, 241-257.
- [105] Patz, R. J. e Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24, 146-178.
- [106] Prentice, R. L. (1976). A Generalization of the Probit and Logit Methods for Dose-Response Curves. *Biometrika*, 32,761-768.
- [107] Riddle, A., Blais, M., Hess U. (2002). A Multi-Group Investigation of the CES-D's Measurement Structure Across Adolescents, Young Adults and Middle-Aged Adults. Scientific Series 2002s-36. CIRANO.
- [108] Roberts, C., P.(2001) *The Bayesian Choice: from decision-theoretic foundations to computational implementation*. 2nd ed. New york: Springer-Verlag.
- [109] Rupp, A., Dey, D. K., e Zumbo, B. (2004) (in press). To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian Methodology To Item Response Modeling. *Structural Equations Modeling*.
- [110] Sahu, S. K. (2002). Bayesian Estimation and Model Choice in Item Response Models. *Journal of Statistical Computation and Simulation*, 72, 217-232.
- [111] Sahu, S. K., Dey, D.E. e Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, Vol. 29,217-232.

- [112] Samejima, F. (1997). Departure from normal assumptions: a promise for future psychometrics with substantive mathematical modeling. *Psychometrika*, 62, 4, 471-493.
- [113] Samejima, F. (2000). Logistic positive exponent family of models: virtue of asymmetric item characteristics curves. *Psychometrika*, 65(3), 319-335.
- [114] Segall, D. (2003). Confirmatory Item Factor Analysis using Markov Chain Monte Carlo Estimation with Applications to Online Calibration in CAT. Pre-print.
- [115] Spiegelhalter, D. J., Thomas, A., Best, N. G. e Gilks, W.R.(1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambrigde, UK: University of Cambride.
- [116] Spiegelhalter D., Thomas, A., Best N. e Lunn, D. (2003). *WinBUGS version 1.4* [Computer program]. Imperial College & Medical Research Council Bioestatics Unit. Institute of Public Health, Cambridge University.
- [117] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. e Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Royal Statistical Society, Series B*, 64(3), 583-639.
- [118] Stern, H. e Sinharay, S. (2004). Model Checking and Model Diagnostics. Preprint.
- [119] Stukel, T. (1988). Generalized Logistic Models. *Journal of the American Statistical Association*, 83, 426-431.
- [120] Swaminathan, H. e Gifford, J. A. (1982). Bayesian Estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175 - 192.
- [121] Swaminathan, H. e Gifford, J. A. (1985). Bayesian Estimation in the two-parameter logistic model. *Psychometrika*, 50, 349 - 364.
- [122] Swaminathan, H. & Gifford, J. A. (1986). Bayesian Estimation in the three-parameter logistic model. *Psychometrika*, 51, 589 - 601.
- [123] Tanner, M. A. (1996). *Tools for Statistical Inference*. Springer-Verlag: Heidelberg.
- [124] Tavares, H. R. *Teoria da Resposta ao Item para Dados Longitudinais*. Tese de Doutorado. São Paulo: IME/USP.
- [125] Taylor, J. e Siqueira, A. (1996). The cost of Adding Parameters to a Model. *Journal of the Royal Statistical Society*, 58, 593-607.
- [126] Tsutakawa, R. K. e Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika*, 51, 251-267.

- [127] Tsutakawa, R. M. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics*, 9, 263-276.
- [128] UMC (2001). *Análisis de items de las pruebas Crecer 1998. Resultados de lógico matemática en sexto grado de primaria*. Boletín 13. Unit of Mensuration of Educational Quality. Ministry of Education. Peru. Available from <http://www.minedu.gob.pe/umc>
- [129] Valle, R. C. (1999). *Teoria da Resposta ao Item*. Dissertação de Mestrado. São Paulo : IME/USP.
- [130] Van der Linden, W.J. e Hambleton, R. K. (1997). *Handbook of Modern Item response Theory*. New York: Springer-Verlag.
- [131] Vianna, H. (2003). Avaliações Nacionais em Larga Escala: análise e propostas. *Estudos em avaliação Educacional*, 27, 41-76.
- [132] Vidal, I. (2003). Bayesian Model Comparison for Skew-Elliptical Regression Models. Tese de Doutorado. PUC-Chile.
- [133] Vidal, I., Iglesias, P., Branco, M.D. e Arellano-Valle, R.B.(2005). Bayesian Sensitivity Analysis and Model Comparison for Skew Elliptical Models. *Journal of Statistical Planning and Inference*, accepted.
- [134] Wang, X., Bradlow, E. T. e Wainer, H. (2003). A General Bayesian Model for Testlets: Theory and Applications. *Applied Psychological Measurement*, 26, 109-128.
- [135] Yan D., Mislevy, R.J. e Almond, R.G. (2003). Design and analysis in a cognitive assesment. ETS Research Report 03-32, Educational Testing Service, Princeton, N.J.
- [136] Zaidler, T.I., Heimberg, R.G., Fresco, D.M., Schneier, F.R. e Liebowitz, M.R. (2003). Evaluation of the Clinical Global Impression Scale among individual with social anxiety disorder. *Psychological Medicine*, 33, 611-622.
- [137] Zellner, A. e Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, 25, 365-393.