**Regressão binária bayesiana
com o uso de variáveis auxiliares**

Rafael Bráz Azevedo Farias

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

Área de Concentração: **Estatística**
Orientadora: **Profa. Dra. Márcia D’Elia Branco**

*Durante a elaboração deste trabalho o autor
recebeu apoio financeiro do CNPq*

– São Paulo, Abril de 2007 –

**Regressão binária bayesiana
com o uso de variáveis auxiliares**

Este exemplar corresponde à redação
final da dissertação devidamente corrigida e
defendida por Rafael Bráz Azevedo Farias
e aprovada pela comissão julgadora.

São Paulo, 27 de Abril de 2007

Banca Examinadora:

- Prof. Dra. Márcia D'Elia Branco (Orientadora) - IME/USP
- Prof. Dra. Mônica Carneiro Sandoval - IME/USP
- Prof. Dr. Luis Aparecido Milan -UFSCar

“Para aqueles que amam como se jamais pudessem se machucar...”

Caio Azevedo

“A razão cardeal de toda a superioridade humana é sem dúvida a vontade. O poder nasce do querer. Sempre que o homem aplique a veemência e perseverante energia de sua alma a um fim, ele vencerá os obstáculos, e se não atingir o alvo, fará pelo menos coisas admiráveis.”

José de Alencar

“Não sabendo que era impossível, ele foi lá e fez.”

Jean Cocteau

AGRADECIMENTOS

Aos meus pais, Afonso e Fátima, pelo amor, carinho, e por serem para mim exemplos de vida. Muito obrigado por tudo!

À minha irmã Daniele pelo apoio e brincadeiras de infância. À minha priminha Cinthia por todo amor e por iluminar minha vida de alegrias.

À minha orientadora professora Márcia D'Elia Branco, pela confiança em mim depositada, pela paciência *ad infinitum*, e pela orientação inestimável que recebi ao longo do desenvolvimento deste trabalho.

Aos professores do Departamento de Estatística do IME-USP pelos valiosos ensinamentos recebidos, principalmente à Mônica Sandoval, Júlia Pavan, Chang Chiann, Anatoli Iambartsev, Gilberto Alvarenga, Márcia Branco e Silvia Ferrari pelos cursos ministrados. Aos funcionários da USP pela prontidão em diversos momentos e esclarecimentos prestados.

Aos professores do Departamento de Estatística e Matemática Aplicada da UFC, por me fornecerem uma base sólida na minha caminhada, especialmente ao grande professor João Maurício pelos ensinamentos, conselhos e por sua amizade. Devo citar também o professor João Welliandre pela grande ajuda durante a graduação.

Aos membros da comissão examinadora por disponibilizarem seus tempos avaliando este trabalho e pelas valiosas sugestões e comentários, em especial à professora Mônica Sandoval.

Ao Amigo Alexandre Patriota pela enorme ajuda durante a graduação e mestrado, e por ser como um irmão nos últimos dois anos. Aos amigos de graduação Ênio Lopes, Chagas Almeida, Michel Helcias, Fabienne Rodrigues e Eveliny Barroso, pelas brincadeiras, conversas e pelos inúmeros momentos agradáveis. Aos amigos Juvêncio Nobre, Jacqueline Batista, Caio Azevedo, Iesus Diniz e Márcio Medeiros, pela acolhida durante o primeiro ano de mestrado, e pelos momentos de descontração e aprendizado.

Aos grandes amigos Alexandre, Caio, Juvêncio e Michel por estarem sempre preocupados com meu bem estar e fazerem com que eu me sinta entre irmãos.

Aos colegas do IME-USP pelo companheirismo, pela ajuda nas horas difíceis e pelos momentos de descontração, dentre os quais destaco: Germán, Estevão, Artur, Ivan, Dalton, Marcelo, Ronald, Gilberto, Alessandro, Nubia, Marina, Betsabé, Luciana e Lauren.

Aos amigos Alvaro Diego, Andreia Gouveia e Antonio Lemes, pelas conversas e por me ajudarem a conhecer melhor a cidade de São Paulo.

Aos meus tios Cleide, Milton, Zuleide e Chico Mendes, pelo carinho, compreensão e amizade durante minha infância e adolescência.

Aos meus primos e amigos que estão no Ceará: George, Georcilene, Georgiane, Roberto, Humberto, Alexandre, Fábio, Fabiana, Denize, Luiz, Diego, Laudenir, Daniel, Clebio e Argeu.

A todos os amigos de Fortaleza e São Paulo que de alguma forma contribuíram para esse momento.

Ao CNPq pelo apoio financeiro.

Acima de tudo, agradeço a Deus por seu amor e sua proteção.

RESUMO

A inferência Bayesiana está cada vez mais dependente de algoritmos de simulação estocástica, e sua eficiência está diretamente relacionada à eficiência do algoritmo considerado. Uma prática bastante utilizada é a introdução de variáveis auxiliares para obtenção de formas conhecidas para as distribuições *a posteriori* condicionais completas, as quais facilitam a implementação do amostrador de Gibbs. No entanto, a introdução dessas variáveis pode produzir algoritmos onde os valores simulados são fortemente correlacionados, fato esse que prejudica a convergência. O agrupamento das quantidades desconhecidas em blocos, de tal maneira que seja viável a simulação conjunta destas quantidades, é uma alternativa para redução da autocorrelação, e portanto, ajuda a melhorar a eficiência do procedimento de simulação. Neste trabalho, apresentamos propostas de simulação em blocos no contexto de modelos de regressão binária com o uso de variáveis auxiliares. Três classes de funções de ligação são consideradas: probito, logito e probito-assimétrico. Para as duas primeiras apresentamos e implementamos as propostas de atualização conjunta feitas por Holmes e Held (2006). Para a ligação probito-assimétrico propomos quatro diferentes maneiras de construir os blocos, e comparamos estes algoritmos através de duas medidas de eficiência (distância média Euclidiana entre atualizações e tamanho efetivo da amostra). Concluimos que os algoritmos propostos são mais eficientes que o convencional (sem blocos), sendo que um deles proporcionou ganho superior a 160% no tamanho efetivo da amostra. Além disso, discutimos uma etapa bastante importante da modelagem, denominada análise de resíduos. Nesta parte adaptamos e implementamos os resíduos propostos para a ligação probito para os modelos logístico e probito-assimétrico. Finalmente, utilizamos os resíduos propostos para verificar a presença de observações discrepantes em um conjunto de dados simulados.

ABSTRACT

The Bayesian inference is getting more and more dependent of stochastic simulation algorithms, and its efficiency is directly related with the efficiency of the considered algorithm. The introduction of auxiliary variables is a technique widely used for attainment of the full conditional distributions, which facilitate the implementation of the Gibbs sampling. However, the introduction of these auxiliary variables can produce algorithms with simulated values highly correlated, this fact harms the convergence. The grouping of the unknown quantities in blocks, in such way that the joint simulation of this quantities is possible, is an alternative for reduction of the autocorrelation, and therefore, improves the efficiency of the simulation procedure. In this work, we present proposals of simulation using the Gibbs block sampler in the context of binary response regression models using auxiliary variables. Three class of links are considered: probit, logit and skew-probit. For the two first we present and implement the scheme of joint update proposed by Holmes and Held (2006). For the skew-probit, we consider four different ways to construct the blocks, and compare these algorithms through two measures of efficiency (the average Euclidean update distance between interactions and effective sample size). We conclude that the considered algorithms are more efficient than the conventional (without blocks), where one of these leading to around 160% improvement in the effective sample size. Moreover, we discuss one important stage of the modelling, called residual analysis. In this part we adapt and implement residuals considered in the probit model for the logistic and skew-probit models. For a simulated data set we detect the presence of outlier used the residuals proposed here for the different models.

Conteúdo

Lista de Tabelas	iv
Lista de Figuras	v
1 Introdução	1
1.1 Regressão binária	3
1.2 Modelo bayesiano	5
1.3 Objetivos e organização da dissertação	7
2 Modelos simétricos e o uso de variáveis auxiliares	9
2.1 Regressão probito	11
2.2 Regressão logística	14
3 Modelo probito-assimétrico com o uso de variáveis auxiliares	18
3.1 Regressão probito-assimétrico	20
3.2 Algoritmos de simulação	22
3.2.1 Atualização conjunta de $\{z, \beta\}$	23
3.2.2 Atualização conjunta de $\{z, w\}$	24
3.2.3 Atualização conjunta de $\{z, \lambda\}$	25
3.2.4 Atualização conjunta de $\{z, \beta, \lambda\}$	26
3.3 Comparação dos algoritmos	28
3.3.1 Medidas de eficiência	28
3.3.2 Análise de eficiência, com λ conhecido	30
3.3.3 Análise de eficiência, com λ desconhecido	32

4	Análise de resíduos	35
4.1	Resíduos binários bayesianos	35
4.2	Resíduos latentes	36
4.3	Aplicação	39
5	Considerações finais	52
A	Pseudo-Códigos	54
A.1	Procedimento para amostragem no modelo probito	54
A.1.1	Algoritmo HH	54
A.2	Procedimentos para amostragem no modelo logístico	55
A.2.1	Algoritmo I	55
A.2.2	Algoritmo II	57
A.2.3	Procedimento para amostrar de $\delta \sim \pi(\delta \mathbf{z}, \boldsymbol{\beta})$	57
A.3	Procedimentos para amostragem no modelo probito-assimétrico	60
A.3.1	Algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$	60
A.3.2	Algoritmo $HH(\mathbf{z}, \mathbf{w})$	61
A.3.3	Algoritmo $HH(\mathbf{z}, \lambda)$	63
A.3.4	Algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$	64
B	Método para amostrar da variável misturadora no modelo logístico	66
C	Provas das distribuições do Capítulo 3	70
D	Algumas distribuições utilizadas	83
D.1	Distribuição normal-assimétrica	83
D.2	Distribuição gaussiana-inversa	85
D.3	Distribuição gaussiana-inversa generalizada	86
D.4	Distribuição Kolmogorov-Smirnov	87
	Referências Bibliográficas	88

Lista de Tabelas

3.1	Valores do tempo de sistema (CPU), em segundos, e das medidas <i>ESS</i> e <i>Dist</i> para os diferentes algoritmos	32
3.2	Valores de <i>Dist</i> para os diferentes algoritmos	33
3.3	Valores do tempo de sistema (CPU), em segundos, e <i>ESS</i> para os diferentes algoritmos	34
4.1	Conjunto de dados simulados	41
4.2	Estatísticas <i>a posteriori</i> dos parâmetros regressores para os modelos ajustados	41
4.3	Observações e probabilidades de pontos discrepantes no modelo probito . .	49
4.4	Observações e probabilidades de pontos discrepantes no modelo logito . . .	50
4.5	Observações e probabilidades de pontos discrepantes no modelo probito-assimétrico	51

Lista de Figuras

4.1	Boxplots das distribuições <i>a posteriori</i> dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo probito	42
4.2	Boxplots das distribuições <i>a posteriori</i> dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo logito	43
4.3	Boxplots das distribuições <i>a posteriori</i> dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo probito-assimétrico	43
4.4	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o ajuste do modelo probito	45
4.5	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo logito	45
4.6	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo probito-assimétrico	46
4.7	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\epsilon_i = z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i)$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo probito-assimétrico	47
4.8	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\tau_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta})/\sqrt{\delta_i}$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo logito	48
4.9	Boxplots das distribuições <i>a posteriori</i> dos resíduos latentes $\zeta_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2/(1 + \lambda^2)$ contra as probabilidades ajustadas $\mathbb{E}(p_i \mathbf{y})$ para o modelo probito-assimétrico	48
B.1	Curva da taxa de aceitação estimada para os valores de r_i^2	69

Capítulo 1

Introdução

Em várias situações nos deparamos com variáveis resposta de natureza binária, isso ocorre principalmente em ciências sociais, biológicas e econômicas. A abordagem clássica de modelos de regressão geralmente utiliza a teoria de máxima verossimilhança e as inferências são baseadas na teoria assintótica e, portanto, a precisão das inferências para pequenas amostras é questionável. Na abordagem Bayesiana, as inferências são baseadas nas distribuições *a posteriori* das quantidades de interesse. Infelizmente, em alguns casos, a obtenção das distribuições *a posteriori* é trabalhosa ou exige um custo computacional bastante elevado. A introdução de variáveis adicionais que não estavam presentes na formulação inicial do problema, visa facilitar a construção de algoritmos de simulação das distribuições *a posteriori*.

Albert e Chib (1993) introduziram o uso de variáveis auxiliares em regressão binária bayesiana com a ligação proibito. Neste caso, com a escolha de distribuições *a priori* adequadas, podemos construir um algoritmo de simulação (Amostrador de Gibbs) relativamente simples para gerar uma amostra da distribuição *a posteriori* de interesse. Entretanto, devido à forte correlação *a posteriori* entre os coeficientes regressores e as variáveis auxiliares, este algoritmo não é muito eficiente. Holmes e Held (2006) propuseram um esquema mais eficiente de simulação usando a idéia de “blocos”. A diferença dos algoritmos encontra-se no fato do primeiro, denotado aqui por AC, simular da distribuição *a posteriori* condicional das variáveis auxiliares dado os parâmetros regressores do modelo, e o segundo, denotado aqui por HH, simular da distribuição marginal *a posteriori* das variáveis auxiliares. Este último permite a atualização conjunta dos coeficientes regressores e das variáveis auxiliares.

Um dos modelos comumente utilizado em regressão binária é o modelo logito, ele permite uma simples interpretação dos parâmetros regressores. Neste caso, obtemos uma fácil representação da razão de chances, o que ajuda na interpretação dos resultados. Albert e Chib

(1993) apresentam a ligação t -Student utilizando-se do fato dela poder ser representada como mistura de normais em relação ao parâmetro de escala (Andrews e Mallows, 1974), e propõem o uso desta ligação com 8 graus de liberdade como uma aproximação do modelo logístico. Chen e Dey (1998) representam o modelo logístico de forma exata utilizando a distribuição de Kolmogorov-Smirnov assintótica (KS) (veja Apêndice D) como distribuição de mistura de escala. A distribuição KS é a expansão de uma série infinita, o que dificulta sua simulação. Assim, os autores utilizam técnicas de aproximações numéricas e o esquema de amostragem de Metropolis para a implementação de um algoritmo de simulação. Holmes e Held (2006) descrevem uma maneira alternativa de simular desta distribuição, utilizando apenas o algoritmo de rejeição. Além disso, apresentam dois procedimentos de simulação utilizando blocos, os quais são mais eficientes que o convencional (sem blocos).

Os modelos probito e logito são exemplos de modelos com funções de ligação simétricas, entretanto, estas funções nem sempre fornecem bons ajustes. Isso ocorre quando a probabilidade de uma dada resposta binária se aproxima de um com uma taxa diferente do que aproxima-se de zero. Neste caso, o melhor ajuste é dado com o uso de ligações assimétricas. Prentice (1976) introduziu uma ligação bi-paramétrica que abrange os modelos logito, probito e algumas ligações assimétricas como casos limites. Aranda-Ordaz (1981) propôs uma ligação uni-paramétrica que tem como casos particulares os modelos logito e complemento log-log. Stukel (1988) definiu uma classe de ligações bi-paramétrica que generaliza o modelo logístico e contém algumas ligações assimétricas como casos particulares. Mais recentemente, motivados pela idéia de variáveis auxiliares apresentada em Albert e Chib (1993), Chen, Dey e Shao (1999) definem uma nova classe de funções de ligação assimétricas para modelos de dados binários, denominada *skew-probit* (ou probito-assimétrico), e tem como caso particular o modelo probito. Bazán, Branco e Bolfarine (2005) utilizaram uma outra função de ligação, também denominada probito-assimétrico, baseada na distribuição normal-assimétrica padrão de Azzalini (1985). A relação entre essas duas ligações é discutida em Bazán, Bolfarine e Branco (2006). Nesta dissertação trabalhamos com a ligação probito-assimétrico proposta por Chen *et al* (1999) e construímos e implementamos quatro diferentes algoritmos de simulação utilizando as idéias de blocos.

As técnicas de diagnóstico são ferramentas indispensáveis para verificação da adequabilidade do ajuste, em particular, os resíduos são utilizados para detectar a presença de

observações discrepantes e para verificar possíveis afastamentos das suposições feitas para o modelo. Albert e Chib (1995) propuseram dois tipos de resíduos em modelos de regressão binária. O primeiro baseia-se na comparação entre a variável observada e sua esperança, e o segundo é baseado na noção de variável auxiliar. Neste trabalho adaptamos e implementamos resíduos propostos no modelo probito para os modelos logístico e probito-assimétrico. Além disso, descrevemos técnicas de detecção de observações discrepantes utilizando os resíduos nos diferentes modelos.

1.1 Regressão binária

Variáveis binárias são variáveis categóricas que podem assumir somente dois estados, denotados, por conveniência, por 0 (“fracasso”) e 1 (“sucesso”). Sucesso significa a ocorrência do evento de interesse. Essas variáveis estão comumente associadas a outras variáveis, que podem ser contínuas, discretas ou categóricas. Considerando que a probabilidade de sucesso possa ser explicada por estas outras variáveis, denominadas variáveis explicativas ou covariáveis, o modelo de regressão binária estabelece a forma funcional desta relação.

Considere $\mathbf{y} = (y_1, \dots, y_n)^\top$ um conjunto de variáveis binárias (0 – 1), sendo $y_1, \dots, y_n$ variáveis aleatórias independentes. Considere também $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ um conjunto de quantidades previamente fixadas associadas a y_i , onde x_{i1} pode ser igual a 1 (que corresponde ao intercepto). O modelo de regressão binária com respostas independentes é definido como

$$p_i = \mathbb{P}(y_i = 1) = F(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad (1.1)$$

sendo F^{-1} uma função que lineariza a relação entre a probabilidade de sucesso e as covariáveis, e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é um vetor de dimensão p de coeficientes regressores. Na teoria de modelos lineares generalizados (MLG), a função F^{-1} é chamada de função de ligação.

Em MLG, o inverso da função de ligação é uma função monótona e diferenciável. Deste modo, assume-se tipicamente que F é uma função de distribuição acumulada (fda) de uma variável aleatória com suporte nos reais, que pode depender de parâmetros adicionais.

Neste contexto, três tipos de funções de ligação podem ser consideradas: (i) função de ligação fixa; (ii) função de ligação paramétrica; e (iii) função de ligação não paramétrica.

As funções de ligação fixas comumente utilizadas são: probito [$F(u) = \Phi(u)$, sendo $\Phi(\cdot)$ a fda de uma normal padrão], logito [$F(u) = \frac{\exp(u)}{1+\exp(u)}$], complemento log-log [$F(u) = 1 + \exp\{-\exp(u)\}$] e t -Student com o número de graus de liberdade conhecido.

O procedimento de estimação clássico para modelos binários, em geral, é baseado na função de verossimilhança. Neste caso, quando a matriz de delineamento $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ tem posto p , a função de verossimilhança é dada por

$$L = \prod_{i=1}^n F(\mathbf{x}_i^\top \boldsymbol{\beta})^{y_i} \left[1 - F(\mathbf{x}_i^\top \boldsymbol{\beta})\right]^{1-y_i}, \quad (1.2)$$

sendo F uma fda que pode depender de parâmetros desconhecidos.

Considerando, por simplicidade, que a função de ligação é fixa, isto é, F não tem parâmetros desconhecidos, obtemos o correspondente vetor escore

$$U(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top (\mathbf{y} - \mathbf{p}),$$

onde $\mathbf{p} = (p_1, \dots, p_n)$ denota o vetor de probabilidades de sucesso, sendo $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta})$. A matriz de informação de Fisher associada é definida por

$$\mathbf{I}(\boldsymbol{\beta}) = \mathbb{E} \left[-\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right] = \mathbf{X}^\top \mathbf{D} \mathbf{X}, \quad (1.3)$$

em que $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$, com $d_i = p_i(1 - p_i)$.

Os estimadores de máxima verossimilhança (EMV) deste modelo são obtidos através das soluções das equações de log-verossimilhanças $\left(\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}\right)$. Se o logaritmo da função de verossimilhança $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ é côncavo, então o EMV é único quando existe pelo menos um $\hat{\boldsymbol{\beta}}$ dentro do conjunto de parâmetros admissíveis que atinge o máximo global ou local da função $l(\boldsymbol{\beta})$. Para obtenção das estimativas de máxima verossimilhança, usa-se métodos iterativos, em particular, o método *Escore de Fisher*, dado por

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{I}(\boldsymbol{\beta}^{(m)})^{-1} U(\boldsymbol{\beta}^{(m)}), \quad (1.4)$$

onde $\boldsymbol{\beta}^{(m)}$ denota o valor de $\boldsymbol{\beta}$ no passo m do método *Escore de Fisher*. Tem-se, sob condições gerais de regularidade (vide, Sen e Singer, 1993, Cap. 7), que o EMV $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ é um estimador eficiente e consistente, e que

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\beta})), \quad \text{quando } n \longrightarrow \infty,$$

em que $\xrightarrow{\mathcal{D}}$ significa convergência em distribuição e $\mathbf{I}(\boldsymbol{\beta})$ é definida em (1.3).

1.2 Modelo bayesiano

Na metodologia bayesiana, assumimos que o vetor de parâmetros desconhecidos tem um modelo probabilístico associado, denominada distribuição *a priori*. O modelo bayesiano de regressão binária considerado neste trabalho é dado por

$$\begin{aligned} y_i &\sim \text{Bernoulli}(F(\eta_i)) \\ \eta_i &= \mathbf{x}_i^\top \boldsymbol{\beta} \\ \boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}), \end{aligned} \tag{1.5}$$

onde $y_i \in \{0, 1\}$, $i = 1, \dots, n$, é uma variável com resposta binária associada a p covariáveis $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$, F^{-1} é a função de ligação, η_i denota o i -ésimo preditor linear e $\boldsymbol{\beta}$ representa um vetor coluna ($p \times 1$) de coeficientes regressores com alguma distribuição *a priori* $\pi(\boldsymbol{\beta})$. Se existir parâmetros desconhecidos associados à função de ligação, então devemos definir também uma distribuição *a priori* para estes parâmetros.

A função de verossimilhança associada ao modelo (1.5) é a mesma definida em (1.2). Neste caso, escrevendo $\mathbf{y} = (y_1, \dots, y_n)^\top$ e assumindo função de ligação fixa, a distribuição dos parâmetros, atualizada pelos dados, utilizando o teorema de Bayes é denominada distribuição *a posteriori* e é dada por

$$\pi(\boldsymbol{\beta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\beta}) \prod_{i=1}^n F(\eta_i)^{y_i} [1 - F(\eta_i)]^{1-y_i}}{\int \prod_{i=1}^n F(\eta_i)^{y_i} [1 - F(\eta_i)]^{1-y_i} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}}. \tag{1.6}$$

Entretanto, essa distribuição não tem forma conhecida, pois não existe uma expressão fechada para a constante normalizadora (denominador), o que torna essa distribuição analiticamente não tratável. Uma alternativa é o uso de métodos de simulação para obtenção de uma amostra dessa distribuição *a posteriori*, em particular, os métodos de Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo*: MCMC).

Os métodos MCMC são um conjunto de técnicas de simulação estocástica para obtenção de amostras de uma distribuição comumente desconhecida. Esses métodos são baseados na construção de cadeias de Markov cuja distribuição de interesse é a distribuição invariante (estacionária). No contexto de inferência bayesiana, a distribuição estacionária é a distribuição *a posteriori*.

Entre os algoritmos MCMC mais conhecidos estão o algoritmo de Metropolis-Hastings, o algoritmo de rejeição adaptativa e o algoritmo Amostrador de Gibbs. Para implementação deste último, devemos saber amostrar das distribuições condicionais completas. Todos estes algoritmos são descritos em Gamerman e Lopes (2006).

A seguir descrevemos o algoritmo Amostrador de Gibbs. Considere $\pi(\boldsymbol{\theta})$ a distribuição de interesse, onde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)^\top$, e cada componente $\boldsymbol{\theta}_i$ pode ser um escalar ou um vetor. Considere também que é possível amostrar das distribuições condicionais completas $\pi_i(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{-i})$, $i = 1, \dots, p$, sendo $\boldsymbol{\theta}_{-i}$ o vetor $\boldsymbol{\theta}$ com o i -ésimo componente removido.

O algoritmo de Gibbs permite uma forma de geração baseada em sucessivas gerações das distribuições condicionais completas. O algoritmo segue o seguinte esquema:

- i) inicializa-se o contador de iterações da cadeia, $j = 1$, e arbitra-se valores iniciais para $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_p^{(0)})^\top$;
- ii) obtem-se um valor para $\boldsymbol{\theta}^{(j)}$ a partir de $\boldsymbol{\theta}^{(j-1)}$ usando sucessivas gerações de valores

$$\begin{aligned} \boldsymbol{\theta}_1^{(j)} &\sim \pi_1 \left(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(j-1)}, \dots, \boldsymbol{\theta}_p^{(j-1)} \right) \\ \boldsymbol{\theta}_2^{(j)} &\sim \pi_2 \left(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j-1)}, \dots, \boldsymbol{\theta}_p^{(j-1)} \right) \\ &\vdots \\ \boldsymbol{\theta}_p^{(j)} &\sim \pi_p \left(\boldsymbol{\theta}_p | \boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, \dots, \boldsymbol{\theta}_{p-1}^{(j)} \right); \end{aligned}$$

- iii) altera-se o contador j para $j + 1$ e retorna-se a (ii) até a convergência.

Uma prática bastante utilizada é a introdução de variáveis auxiliares para obtenção de formas conhecidas das distribuições condicionais completas, o que facilita a implementação do Amostrador de Gibbs. A introdução dessas variáveis visa à obtenção de distribuições condicionais completas mais fáceis de serem simuladas.

Albert e Chib (1993) introduziram variáveis auxiliares ao modelo de regressão binária. A adição dessas variáveis foi motivada pela idéia de variável de tolerância no contexto biológico, e seu uso, combinado às ferramentas do Amostrador de Gibbs, permite uma simples geração de amostras da distribuição *a posteriori*.

O modelo (1.5) com função de ligação simétrica pode ser representado utilizando variáveis auxiliares da seguinte forma

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \\ z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \\ \epsilon_i &\sim F \\ \boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}), \end{aligned} \tag{1.7}$$

sendo $\mathbf{z} = (z_1, \dots, z_n)^\top$ o vetor com n variáveis auxiliares e F uma fda simétrica.

Como F em (1.7) é uma distribuição simétrica em torno do zero, temos que $F(z) = 1 - F(-z)$, e a probabilidade de sucesso no modelo (1.7) condicionada à $\boldsymbol{\beta}$ é dada por

$$P(y_i = 1|\boldsymbol{\beta}) = P(z_i > 0|\boldsymbol{\beta}) = P(\epsilon_i > -\mathbf{x}_i^\top \boldsymbol{\beta}|\boldsymbol{\beta}) = 1 - F(-\mathbf{x}_i^\top \boldsymbol{\beta}) = F(\mathbf{x}_i^\top \boldsymbol{\beta}).$$

Logo, $y_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$, sendo $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta})$ a probabilidade do i -ésimo sucesso.

1.3 Objetivos e organização da dissertação

Os objetivos deste trabalho estão divididos em três partes. Primeiramente apresentamos e implementamos os algoritmos de simulação propostos por Holmes e Held (2006) para os modelos de regressão probito e logístico. Em seguida construímos e implementamos quatro algoritmos de atualização conjunta utilizando variáveis auxiliares para o modelo probito-assimétrico proposto por Chen *et al* (1999). Esses algoritmos generalizam o esquema proposto por Holmes e Held (2006) para a ligação probito. Finalmente, descrevemos os resíduos bayesianos propostos por Albert e Chib (1993), e algumas técnicas de detecção de observações discrepantes em regressão binária Bayesiana para os modelos simétricos. Adaptamos estes resíduos para os modelos logístico e probito-assimétrico e propomos um resíduo baseado em variáveis auxiliares que generaliza o resíduo latente proposto por Albert e Chib (1993).

Esta dissertação é dividida em cinco capítulos e quatro apêndices. No segundo capítulo apresentamos o modelo de regressão simétrico utilizando variáveis auxiliares, e mostramos como obter a distribuição *a posteriori* conjunta dos coeficientes regressores e das variáveis

auxiliares nos modelos probito e logístico. No terceiro capítulo apresentamos a regressão probito-assimétrico utilizando dois conjuntos de variáveis auxiliares, e propomos quatro algoritmos para obtenção da distribuição *a posteriori* conjunta dos parâmetros desconhecidos e das variáveis auxiliares. Finalmente, fazemos um estudo de eficiência dos algoritmos. No quarto capítulo utilizamos as variáveis auxiliares para definir resíduos latentes, apresentamos algumas técnicas de detecção de observações discrepantes em modelos bayesianos de regressão binária, e aplicamos os resíduos em um conjunto de dados simulados. No quinto capítulo apresentamos as conclusões obtidas neste trabalho e perspectivas futuras de pesquisa.

No Apêndice A apresentamos os pseudo-códigos para implementação computacional de todos os algoritmos. No Apêndice B mostramos como gerar da distribuição *a posteriori* da variável de mistura no modelo logístico. No Apêndice C apresentamos as demonstrações para obtenção das distribuições *a posteriori* do modelo probito-assimétrico e as relações matriciais utilizadas. No Apêndice D apresentamos algumas distribuições de probabilidades utilizadas que não são comumente encontradas em textos básicos de probabilidade.

Capítulo 2

Modelos simétricos e o uso de variáveis auxiliares

Os modelos comumente utilizados em regressão binária são modelos obtidos através de funções de distribuições acumuladas de distribuições simétricas. Os mais populares são os modelos probito e logito, adequados quando não há indícios de que a probabilidade de sucesso cresce numa taxa diferente que decresce.

Um ampla classe de distribuições contínuas, unimodais e simétricas com suporte nos reais são representadas estocasticamente por $z = \sqrt{\delta}z_0$, onde z_0 tem distribuição normal padrão e δ é uma variável aleatória não negativa. Estas distribuições são chamadas de esféricas representáveis (vide, por exemplo, Branco e Arellano-Valle, 2004). Note que a distribuição de z condicionada à δ é normal com média zero e variância δ . Por esta razão, esta classe também é chamada de classe das distribuições obtidas através de misturas de normais em relação ao parâmetro de escala (Andrews e Mallows, 1974). Mais precisamente, essa classe é obtida quando consideramos que o parâmetro de escala de uma distribuição normal tem distribuição de probabilidade no intervalo $[0, \infty)$. A classe é dada por

$$\mathcal{F}_S = \left\{ F(\cdot) = \int_{[0, \infty)} \Phi\left(\frac{\cdot}{\delta}\right) dG(\delta), \text{ } G \text{ é uma fda no intervalo } [0, \infty) \right\}, \quad (2.1)$$

e contém várias distribuições conhecidas, com caudas mais pesadas ou mais leves que a distribuição normal. Temos ainda que, todos os membros desta classe são distribuições simétricas quando δ tem uma distribuição contínua. Alguns exemplos são as distribuições t -Student, logística, exponencial dupla, exponencial potência e Cauchy.

Os modelos de misturas de normais em relação ao parâmetro de escala foram descritos no contexto de regressão binária, por exemplo, nos trabalhos de Albert e Chib (1993), Chen e Dey (1998) e Basu e Mukhopadhyay (2000). Albert e Chib (1993) descreveram os modelos probito e t -Student; Chen e Dey (1998) e Basu e Mukhopadhyay (2000) descreveram os modelos probito, t -Student, logito e exponencial potência.

Considerando $F \in \mathcal{F}_S$ no modelo (1.1), obtemos a seguinte representação para regressão binária

$$P(y_i = 1) = F(\mathbf{x}_i^\top \boldsymbol{\beta}) = \int_{[0, \infty)} \Phi\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\delta}\right) dG(\delta). \quad (2.2)$$

Alguns casos particulares são:

- Modelo probito: ocorre quando δ tem distribuição degenerada no ponto 1;
- Modelo t -Student com ν graus de liberdade: ocorre quando δ tem distribuição $\text{gama}(\nu/2, \nu/2)$, que tem função densidade

$$g(\delta) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\nu}{2}\right)^{\nu/2} \delta^{\nu/2-1} \exp\left\{-\frac{\nu}{2}\delta\right\}, \quad \text{com } \delta > 0;$$

- Modelo logito: ocorre quando $\delta = 4\psi^2$ e ψ tem distribuição Kolmogorov-Smirnov assintótica (KS), que tem função densidade dada por

$$g(\delta) = 8 \sum_{n=1}^{\infty} (-1)^{n+1} n^2 \delta \exp(-2n^2 \delta^2), \quad \text{com } \delta > 0.$$

Maiores detalhes sobre a distribuição de Kolmogorov-Smirnov assintótica são apresentados no Apêndice D.

O modelo (2.2) pode ser representado utilizando um vetor de variáveis auxiliares $\mathbf{z} = (z_i, \dots, z_n)^\top$ da seguinte forma,

$$y_i = \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \quad (2.3)$$

com

$$z_i | \boldsymbol{\beta}, \delta \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta(\psi)) \quad \text{e} \quad \psi \sim g(\psi), \quad (2.4)$$

onde $\delta(\psi) > 0$ para todo $\psi > 0$, sendo $\delta(\cdot)$ uma função bijetora e $g(\cdot)$ uma densidade de mistura contínua.

A função de verossimilhança para o modelo apresentado em (2.3)-(2.4) é dada por

$$L(\boldsymbol{\beta} | \mathbf{z}, \mathbf{y}) = \prod_{i=1}^n f(z_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (2.5)$$

onde f denota a função densidade de probabilidade (fdp) associada à fda $F \in \mathcal{F}_S$ e

$\text{Ind}(\mathbf{y}, \mathbf{z}) = \prod_{i=1}^n \{I(y_i = 1)I(z_i > 0) + I(y_i = 0)I(z_i \leq 0)\}$, sendo $I(\cdot)$ uma função indicadora que toma valor 1 se seu argumento for verdadeiro e zero caso contrário. A função $\text{Ind}(\mathbf{y}, \mathbf{z})$ também é uma função indicadora, e toma valor 1 se $(\mathbf{y}, \mathbf{z}) \in R(\mathbf{y}, \mathbf{z})$, onde a região $R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$. A função de verossimilhança (2.5) também é conhecida como função de verossimilhança aumentada.

As distribuições *a posteriori* obtidas a partir do modelo (2.3)-(2.4) são analiticamente mais tratáveis do que as obtidas a partir do modelo usual apresentado em (1.6). Métodos para obtenção de amostras destas distribuições *a posteriori* usando o Amostrador de Gibbs são descritos por diversos autores. Albert e Chib (1993) descrevem como obter as distribuições *a posteriori* no modelo probito e *t*-Student; Chen e Dey (1998) nos modelos de misturas no parâmetro de escala de normais. Holmes e Held (2006) apresentaram uma abordagem alternativa para os modelos probito e logístico. O uso de variáveis auxiliares nos modelos probito e logito serão apresentados com detalhes nas próximas seções.

2.1 Regressão probito

O modelo probito é amplamente utilizado em diversos campos, principalmente em ensaios clínicos. Por exemplo, em modelos de dose-resposta tipicamente estamos interessados em saber a relação entre a probabilidade de um sujeito morrer e a concentração de alguma substância tóxica no sangue. Se a tolerância do sujeito à substância tóxica for assumida ser uma variável normal padrão, temos um modelo de regressão probito. Equivalentemente, obtemos a regressão probito quando assumimos que $F(u) = \Phi(u)$ em (1.5), onde $\Phi(\cdot)$ denota a função de distribuição da normal padrão. Alternativamente, podemos representar a regressão probito bayesiana usando variáveis auxiliares como

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \\ z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, 1) \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\mathbf{b}, \boldsymbol{\nu}), \end{aligned} \tag{2.6}$$

onde y_i é agora condicionalmente determinado pelo sinal da variável auxiliar z_i . A função de verossimilhança do modelo (2.6) é dada em (2.5), sendo f a fdp de uma distribuição

normal padrão. A função de verossimilhança marginal $L(\boldsymbol{\beta}|\mathbf{y})$ é a mesma do modelo (1.5) com $F(\cdot) = \Phi(\cdot)$. A vantagem de trabalhar com a representação (2.6) é que, com uma boa escolha para a distribuição *a priori* de $\boldsymbol{\beta}$, podemos obter uma performance eficiente de simulação. Isto ocorre quando consideramos que $\boldsymbol{\beta}$ tem distribuição *a priori* normal.

Albert e Chib (1993) derivam as distribuições condicionais $\pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{y})$ e $\pi(\boldsymbol{\beta}|\mathbf{z}, \mathbf{y})$ no modelo (2.6). O uso da distribuição *a priori* normal para $\boldsymbol{\beta}$ permite a construção de um algoritmo de Gibbs relativamente simples. A distribuição $\pi(\boldsymbol{\beta}|\mathbf{z}, \mathbf{y})$ é igual a $\pi(\boldsymbol{\beta}|\mathbf{z})$, pois $\mathbf{z}$ determina completamente o valor de $\mathbf{y}$. Essa distribuição é dada por

$$\begin{aligned}\boldsymbol{\beta}|\mathbf{z} &\sim \mathcal{N}_p(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V}(\boldsymbol{\nu}^{-1}\mathbf{b} + \mathbf{X}^\top \mathbf{z}) \\ \mathbf{V} &= (\boldsymbol{\nu}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1},\end{aligned}\tag{2.7}$$

sendo $\mathbf{X} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top, \dots, \mathbf{x}_n^\top)$. Dado $\boldsymbol{\beta}$, os z_i 's são variáveis aleatórias independentes com distribuição normal truncada dada por

$$z_i|y_i, \boldsymbol{\beta} \propto \begin{cases} \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)I(z_i \leq 0) & \text{caso contrário.} \end{cases}\tag{2.8}$$

Podemos usar o método da transformação inversa descrito em Devroye (1986) para amostrar da distribuição normal truncada univariada apresentada em (2.8).

A inclusão de variáveis auxiliares no modelo (2.6) oferece uma estrutura conveniente para o uso do Amostrador de Gibbs. Nesta abordagem vamos realizar amostragens sucessivas das distribuições condicionais (2.7) e (2.8). Entretanto, um problema em potencial está na forte correlação *a posteriori* entre $\boldsymbol{\beta}$ e $\mathbf{z}$, claramente indicada no modelo (2.6). Esta correlação causa uma lenta mistura na cadeia. Sob o modelo (2.6), Holmes e Held (2006) propõem um esquema de simulação mais eficiente do que o esquema proposto por Albert e Chib (1993). A diferença dos algoritmos encontra-se no fato do AC gerar da distribuição *a posteriori* condicional das variáveis auxiliares dado os parâmetros regressores do modelo, e o HH, da distribuição marginal *a posteriori* das variáveis auxiliares. Este último permite uma atualização conjunta dos coeficientes da regressão e das variáveis auxiliares usando a seguinte fatoração

$$\pi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{y}) = \pi(\mathbf{z}|\mathbf{y})\pi(\boldsymbol{\beta}|\mathbf{z}),$$

onde a distribuição $\pi(\boldsymbol{\beta}|\mathbf{z})$ é a mesma dada em (2.7), mas agora $\mathbf{z}$ é atualizado de sua

distribuição marginal *a posteriori*. Holmes e Held (2006) assumem que β tem distribuição *a priori* normal n -variada com vetor de médias zero e encontram a distribuição de $\mathbf{z}$ condicionada à $\mathbf{y}$. Apresentamos aqui um resultado mais geral, onde as médias *a priori* de β não são necessariamente iguais a zero.

Sob o modelo (2.6), a distribuição condicional de $\mathbf{z}$ dado $\mathbf{y}$ é dada por

$$\mathbf{z}|\mathbf{y} \propto \mathcal{N}_n(\mathbf{X}\mathbf{b}, \mathbf{I}_n + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top)\text{Ind}(\mathbf{y}, \mathbf{z}), \quad (2.9)$$

sendo $\mathbf{I}_n$ a matriz identidade de dimensão n e $\text{Ind}(\mathbf{y}, \mathbf{z})$ é a mesma função indicadora apresentada em (2.5).

É complicado amostrar diretamente da distribuição normal n -variada truncada apresentada em (2.9). Contudo, uma forma eficiente de obter uma amostra desta distribuição, é usar um algoritmo de Gibbs. Neste caso, a distribuição condicional de cada componente z_i , $i = 1, \dots, n$, condicionada aos demais, é normal univariada truncada no zero dada por

$$z_i|\mathbf{z}_{-i}, \mathbf{y} \propto \begin{cases} \mathcal{N}(m_i, \nu_i)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(m_i, \nu_i)I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (2.10)$$

onde $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. Os parâmetros dessa distribuição são $m_i = \mathbf{x}_i^\top \mathbf{b} + (1 - h_{ii})^{-1} \sum_{k=1, k \neq i}^n h_{ik}(z_k - \mathbf{x}_k^\top \mathbf{b})$ e $\nu_i = (1 - h_{ii})^{-1}$, sendo $h_{ik} = \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_k$, com $\mathbf{V}$ definida em (2.7). Uma maneira eficiente de calcular os parâmetros de localização é dada por

$$m_i = \mathbf{x}_i^\top \mathbf{B} - \frac{h_{ii}}{1 - h_{ii}}(z_i - \mathbf{x}_i^\top \mathbf{B}),$$

sendo z_i o valor atual da i -ésima observação do vetor $\mathbf{z}$ e $\mathbf{B}$ é dado em (2.7). Para cada atualização de algum z_i , devemos recalculer o vetor $\mathbf{B}$ usando a relação

$$\mathbf{B} = \mathbf{B}^{ant} + \mathbf{s}_i (z_i - z_i^{ant}),$$

onde $\mathbf{B}^{ant}$ e z_i^{ant} são os valores armazenados da atualização anterior de $\mathbf{B}$ e z_i , respectivamente, e $\mathbf{s}_i$ denota o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V} \mathbf{X}^\top$. Esse procedimento é melhor ilustrado no Apêndice A.

Holmes e Held (2006) mostram em um estudo empírico que o algoritmo de atualização conjunta (HH) no modelo probito é mais eficiente o que algoritmo AC.

2.2 Regressão logística

As funções de ligação probito e logito apresentam um comportamento praticamente similar, diferenciando-se apenas nas *caudas*, pois a ligação logística decresce quando a probabilidade de sucesso vai para zero e cresce quando a probabilidade de sucesso vai para um, mais rapidamente que a ligação probito. O uso do modelo de regressão logística tornou-se uma prática comum atualmente. Dentre os motivos, temos a fácil interpretação de seus parâmetros através de razão de chances.

Considere o modelo (2.6), se a suposição de normalidade para ϵ_i for substituída pela suposição de distribuição logística padrão, obtem-se o modelo de regressão logístico. Infelizmente, neste caso, não temos distribuições *a posteriori* conhecidas para atualização dos parâmetros regressores. Entretanto, a distribuição logística pertence à classe de distribuições obtidas através de misturas no parâmetro de escala de normais, definida em (2.2). Assim, adicionando mais um conjunto de variáveis auxiliares $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$, o modelo logito pode ser representado como

$$\begin{aligned}
 y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \\
 z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \\
 \epsilon_i &\sim \mathcal{N}(0, \delta_i) \\
 \delta_i &= (2\psi_i)^2 \\
 \psi_i &\sim KS \\
 \boldsymbol{\beta} &\sim \mathcal{N}_p(\mathbf{b}, \boldsymbol{\nu}),
 \end{aligned} \tag{2.11}$$

sendo ψ_i , $i = 1, \dots, n$, variáveis aleatórias independentes seguindo uma distribuição de Kolmogorov-Smirnov (KS) (veja Apêndice D). A função de verossimilhança do modelo (2.11) é dada em (2.5), sendo f a fdp de uma distribuição logística padrão, e a função de verossimilhança marginal $L(\boldsymbol{\beta}|\mathbf{y})$ é a mesma do modelo (1.5) com função de ligação logística. Uma alternativa para implementação dos métodos MCMC é encontrar as distribuições condicionais completas e utilizar o algoritmo de Gibbs.

Sob o modelo (2.11), a distribuição de $\boldsymbol{\beta}$ condicionada à $\{\mathbf{z}, \boldsymbol{\delta}\}$ é dada por

$$\begin{aligned}\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\delta} &\sim \mathcal{N}_p(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V} \left(\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top \mathbf{W} \mathbf{z} \right) \\ \mathbf{V} &= (\boldsymbol{\nu}^{-1} + \mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \\ \mathbf{W} &= \text{diag}(\delta_1^{-1}, \dots, \delta_n^{-1}).\end{aligned}\tag{2.12}$$

Temos também que a distribuição condicional *a posteriori* de cada z_i dado as demais quantidades é novamente normal truncada, mas agora com variância igual a δ_i , isto é,

$$z_i|y_i, \boldsymbol{\beta}, \delta_i \propto \begin{cases} \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta_i) I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta_i) I(z_i \leq 0) & \text{caso contrário.} \end{cases}\tag{2.13}$$

A distribuição condicional completa $\pi(\boldsymbol{\delta}|\mathbf{z}, \boldsymbol{\beta})$ não tem forma conhecida. Entretanto, mostramos no Apêndice B como gerar uma amostra dessa distribuição utilizando um algoritmo de rejeição. O pseudo-código desse método é apresentado no Apêndice A.

As distribuições condicionais completas apresentadas em (2.12) e (2.13), e o algoritmo para simular de $(\boldsymbol{\delta}|\mathbf{z}, \boldsymbol{\beta})$ apresentado no Apêndice B, nos permite a construção de um algoritmo de Gibbs utilizando as seguintes atualizações sucessivas:

- i) gera-se de $(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\delta})$, em seguida;
- ii) gera-se de $(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{y})$, e por fim;
- iii) gera-se de $(\boldsymbol{\delta}|\mathbf{z}, \boldsymbol{\beta})$ e retorna a (i).

Entretanto, esse esquema de amostragem é mais lento que o caso probito. Isso ocorre porque além de termos adicionado dois conjuntos de variáveis auxiliares, devemos também atualizar a cada iteração a matriz de covariâncias $\mathbf{V}$ definida em (2.12), pois esta é função das variáveis auxiliares $\boldsymbol{\delta}$.

A construção de blocos no modelo logístico é feita de duas maneiras. A primeira segue o mesmo procedimento feito no caso probito e atualiza $\{\mathbf{z}, \boldsymbol{\beta}\}$ conjuntamente utilizando a seguinte fatoração

$$\pi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\delta}) = \pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\delta})\pi(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\delta}).$$

Holmes e Held (2006) assumem que $\mathbf{b} = \mathbf{0}$ em (2.12) e apresentam o pseudo-código para

esse método de atualização conjunta. De um modo geral, para $\mathbf{b} \in \mathbb{R}^p$, a distribuição $\pi(\mathbf{z}|\boldsymbol{\delta}, \mathbf{y})$ é normal n -variada truncada dada por

$$\mathbf{z}|\mathbf{y}, \boldsymbol{\delta} \propto \mathcal{N}_p(\mathbf{X}\mathbf{b}, \mathbf{W}^{-1} + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (2.14)$$

onde $\text{Ind}(\mathbf{y}, \mathbf{z})$ é a mesma função indicadora definida no caso probito. Assim como no modelo probito, usa-se o algoritmo de Gibbs para gerar uma amostra desta distribuição. As distribuições condicionais completas da distribuição (2.14) são dadas por

$$z_i|\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\delta} \propto \begin{cases} \mathcal{N}(m_i, \nu_i)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(m_i, \nu_i)I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (2.15)$$

onde $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. Os parâmetros m_i e ν_i são calculados eficientemente por

$$m_i = \mathbf{x}_i^\top \mathbf{B} - \frac{h_i}{\delta_i - h_i}(z_i - \mathbf{x}_i^\top \mathbf{B}) \quad \text{e} \quad \nu_i = \frac{\delta_i^2}{\delta_i - h_i},$$

onde z_i e δ_i denotam, respectivamente, os valores atuais de z_i e δ_i , $\mathbf{B}$ é dado em (2.12) e h_i é o i -ésimo elemento da diagonal da matriz $\mathbf{H} = \mathbf{X}\mathbf{V}\mathbf{X}^\top$, com $\mathbf{V}$ definido em (2.12). Devemos recalculá-lo a cada atualização de algum z_i através da seguinte relação

$$\mathbf{B} = \mathbf{B}^{ant} + \left(\frac{z_i - z_i^{ant}}{\delta_i} \right) \mathbf{s}_i,$$

sendo $\mathbf{B}^{ant}$ e z_i^{ant} os valores armazenados da atualização anterior de $\mathbf{B}$ e z_i , respectivamente, e $\mathbf{s}_i$ denota o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V}\mathbf{X}^\top$. O próximo passo deste algoritmo é amostrar de $(\boldsymbol{\delta}|\boldsymbol{\beta}, \mathbf{z})$ usando o algoritmo dado no Apêndice A e recalculá-lo $\mathbf{V}$ e $\mathbf{B}$ a cada iteração. Esse algoritmo será denotado por Algoritmo I e seu pseudo-código é apresentado no Apêndice A.

A outra opção é atualizar $\{\mathbf{z}, \boldsymbol{\delta}\}$ conjuntamente utilizando a fatoração abaixo,

$$\pi(\mathbf{z}, \boldsymbol{\delta}|\mathbf{y}, \boldsymbol{\beta}) = \pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta})\pi(\boldsymbol{\delta}|\mathbf{z}, \boldsymbol{\beta}).$$

Neste caso, as distribuições dos z_i 's são logísticas truncadas independentes,

$$z_i|y_i, \boldsymbol{\beta} \propto \begin{cases} Lo(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)I(z_i > 0) & \text{se } y_i = 1, \\ Lo(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (2.16)$$

onde $Lo(a, b)$ denota a função densidade da distribuição logística com parâmetro de localização a e parâmetro de escala b (Devroye, 1986, p. 39). Uma vantagem deste algoritmo,

denotado aqui por Algoritmo II, consiste no fato de ser fácil e eficiente amostrar da distribuição logística truncada usando o método da inversão, isto ocorre porque a fda e sua inversa no modelo logístico tem uma forma analítica simples. Entretanto, observamos que a correlação serial dos parâmetros regressores é menor no Algoritmo I, promovendo uma mistura mais rápida na cadeia. A distribuição de $(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\delta})$ é dada em (2.12) e o pseudo-código para esta abordagem é apresentada no Apêndice A.

Dentre os dois algoritmos de atualizações conjunta apresentados para o modelo logístico, Holmes e Held (2006) recomendam o uso do Algoritmo II devido a simplicidade de seu código.

Capítulo 3

Modelo probito-assimétrico com o uso de variáveis auxiliares

Como já foi dito anteriormente, funções de ligação simétricas podem ser inapropriadas quando a probabilidade de sucesso aproxima-se de zero a uma taxa diferente de quando se aproxima de um. Funções de ligação assimétricas podem ser obtidas considerando que a função de ligação é o inverso da fda de uma distribuição assimétrica. Uma ligação assimétrica bastante utilizada é a complemento log-log (C log-log). Neste caso, a função de ligação é o inverso da fda de uma distribuição Gumbel. Entretanto, não existe um parâmetro que modele essa assimetria. Prentice (1976), Aranda-Ordaz (1981) e Stukel (1988) propuseram funções de ligação paramétricas com parâmetros que controlam a assimetria da ligação. Neste capítulo consideramos a função de ligação assimétrica *skew-probit* (ou probito-assimétrico) proposta por Chen *et al* (1999).

Chen (2004) fez um estudo de simulação para investigar a importância da escolha da função de ligação na predição da variável resposta. Ele considerou dois esquemas de simulação: (i) os dados são gerados de acordo com o modelo probito; e (ii) os dados são gerados de acordo com o modelo C log-log. Em ambas situações, foram ajustados os modelos probito, logito e C log-log. O autor observou que, quando a função de ligação verdadeira é probito, quase não há diferença entre o modelo probito e logito. Entretanto, o modelo C log-log é inadequado. Por outro lado, quando a função de ligação verdadeira é a C log-log, os modelos simétricos foram claramente inadequados. O modelo logito se comportou melhor que o probito neste caso. Isso deve-se ao fato que, para pequenas probabilidades de sucesso, as ligações logística e C log-log encontram-se bastante próximas, decaindo mais rapidamente que a probito. O autor concluiu neste estudo empírico que a escolha da função de ligação é muito importante, e caso seja mal especificada, pode fornecer predições pobres.

Considere a seguinte classe de distribuições para funções de ligação assimétricas pararmétricas

$$\mathcal{F}_A = \left\{ F_\lambda(z) = \int_{[0,\infty)} F(z - \lambda w) dG(w), \lambda \in \mathbb{R} \right\}, \quad (3.1)$$

sendo F uma fda de uma distribuição simétrica em torno do zero com suporte nos reais e G a fda de uma distribuição assimétrica no intervalo $[0, \infty)$. O modelo definido em (3.1) tem algumas propriedades atrativas, tais como: (a) quando $\lambda = 0$ ou G é uma distribuição degenerada, o modelo se reduz ao modelo com função de ligação simétrica; (b) a assimetria da função de ligação pode ser caracterizada por λ e G ; e (c) caudas pesadas e leves para F_λ podem ser obtidas de acordo com a escolha de F , que pode ser uma fda pertencente a classe de mistura no parâmetro de escala de normais, isto é, $F \in \mathcal{F}_S$, com $\mathcal{F}_S$ definida em (2.1). Além dessas, uma propriedade interessante será apresentada na proposição abaixo.

Proposição 3.1 *Considerando F_λ e $F_{-\lambda}$ funções de distribuições pertencentes à classe $\mathcal{F}_A$, temos a seguinte relação entre essas distribuições*

$$F_\lambda(z) = 1 - F_{-\lambda}(-z).$$

$$\textit{Prova: } F_\lambda(z) = \int_{[0,\infty)} F(z - \lambda w) dG(w) = \int_{[0,\infty)} [1 - F(-z + \lambda w)] dG(w) = 1 - F_{-\lambda}(-z). \quad \blacksquare$$

O modelo de regressão binária (1.1), quando o inverso da função de ligação $F_\lambda \in \mathcal{F}_A$, é definido por

$$p_i = \mathbb{P}(y_i = 1) = F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta}) = \int_{[0,\infty)} F(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w) dG(w). \quad (3.2)$$

onde F e G são definidas em (3.1).

Proposição 3.2 *O modelo de regressão binária apresentado em (3.2) é, equivalentemente, definido considerando que*

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \\ z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i^*, \end{aligned} \quad (3.3)$$

sendo

$$\epsilon_i^* = -\lambda w_i + \epsilon_i, \quad \epsilon_i \sim F \quad e \quad w_i \sim G. \quad (3.4)$$

Prova: Note que ϵ_i^* é a representação estocástica considerada em Sahu, Dey e Branco (2003) para construção de variáveis assimétricas. Assim, a variável ϵ_i^* , $i = 1, \dots, n$ tem fda dada por $F_{-\lambda} \in \mathcal{F}_A$. Portanto, a probabilidade de sucesso neste modelo é dada por

$$\mathbb{P}(y_i = 1 | \boldsymbol{\beta}, \lambda) = \mathbb{P}(z_i > 0 | \boldsymbol{\beta}, \lambda) = \mathbb{P}(\epsilon_i^* > -\mathbf{x}_i^\top \boldsymbol{\beta} | \boldsymbol{\beta}, \lambda) = 1 - F_{-\lambda}(-\mathbf{x}_i^\top \boldsymbol{\beta}).$$

Como $F_{-\lambda} \in \mathcal{F}_A$, da Proposição 3.1, temos que $F_\lambda(z) = 1 - F_{-\lambda}(-z)$, logo,

$$\mathbb{P}(y_i = 1 | \boldsymbol{\beta}, \lambda) = F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta})$$

e $y_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$, sendo $p_i = F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta})$ a probabilidade do i -ésimo sucesso. ■

A função de verossimilhança para o modelo (3.3)-(3.4) é dada por

$$L(\boldsymbol{\beta}, \lambda | \mathbf{z}, \mathbf{y}) = \prod_{i=1}^n f_{-\lambda}(z_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (3.5)$$

onde $f_{-\lambda}$ denota a fdp associada a fda $F_{-\lambda} \in \mathcal{F}_A$ e $\text{Ind}(\mathbf{y}, \mathbf{z})$ é a função indicadora na região $R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$, como definida em (2.5).

Um caso particular desse modelo é obtido quando consideramos F uma fda de uma distribuição normal e G a fda de uma distribuição normal positiva (*Half normal*), denotada por $\mathcal{HN}(0, 1)$. A escolha destas distribuições nos retorna o modelo probito-assimétrico.

3.1 Regressão probito-assimétrico

No modelo probito-assimétrico, a função de ligação é o inverso da função de distribuição de uma variável normal-assimétrica (veja Apêndice D). Assim, $F_\lambda(u) = \Phi_{\mathcal{SN}}(u; \mu, \sigma^2, \lambda)$ em (1.1), onde $\Phi_{\mathcal{SN}}(\cdot; \mu, \sigma^2, \lambda)$ denota a fda da distribuição normal-assimétrica definida por Azzalini (1985), com parâmetros de localização μ , de escala σ^2 e de assimetria λ . Se F_λ é a fda de uma distribuição normal-assimétrica pertencente à classe de distribuições

assimétricas $\mathcal{F}_{\mathcal{A}}$ definida em (3.1), então $F_{\lambda}(u) = \Phi_{\mathcal{SN}}(u; 0, 1 + \lambda^2, \lambda) = \Phi_{CDS}(u; \lambda)$, sendo $\Phi_{CDS}(\cdot; \lambda)$ a normal-assimétrica definida em Chen *et al* (1999). Considerando que as distribuições *a priori* de $\mathbf{w} = (w_1, \dots, w_n)^{\top}$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^{\top}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ e λ são independentes, o modelo probito-assimétrico pertencente à classe $\mathcal{F}_{\mathcal{A}}$ pode ser representado por

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário,} \end{cases} \\ z_i &= \mathbf{x}_i^{\top} \boldsymbol{\beta} + \epsilon_i^* \\ \epsilon_i^* &= -\lambda w_i + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, 1) \\ w_i &\sim \mathcal{HN}(0, 1) \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\mathbf{b}, \boldsymbol{\nu}) \\ \lambda &\sim \mathcal{N}(\alpha, \tau). \end{aligned} \tag{3.6}$$

Considerando também independência para os ϵ_i , $i = 1, \dots, n$, a função de verossimilhança marginal $L(\boldsymbol{\beta}, \lambda | \mathbf{y})$ do modelo (3.6) é a mesma do modelo (1.5) com $F(u) = \Phi_{CDS}(u; \lambda)$. Por outro lado, a função de verossimilhança aumentada $L(\boldsymbol{\beta}, \lambda | \mathbf{z}, \mathbf{y})$ é dada em (3.5), com f_{λ} sendo a fdp associada à $\Phi_{CDS}(\cdot; \lambda)$. Portanto, podemos tirar as mesmas conclusões em ambos os modelos.

Uma alternativa ao modelo probito-assimétrico de Chen *et al* (1999) é considerar $F_{\lambda}(u) = \Phi_{\mathcal{SN}}(u; 0, 1, \lambda)$, como proposto em Bazán *et al* (2005). Utilizando o segundo, as variâncias das distribuições *a posteriori* das variáveis auxiliares são funções do parâmetro de assimetria λ , isto dificulta a obtenção da distribuição condicional completa de λ . Neste trabalho, escolheu-se trabalhar com a primeira (CDS), devido a facilidade de implementação do Amostrador de Gibbs.

Sob o modelo (3.6), a distribuição condicional completa de $\boldsymbol{\beta}$ ainda é normal e dada por

$$\begin{aligned} \boldsymbol{\beta} | \mathbf{z}, \mathbf{w}, \lambda &\sim \mathcal{N}_p(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V} \left[\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^{\top} (\mathbf{z} + \lambda \mathbf{w}) \right] \\ \mathbf{V} &= (\boldsymbol{\nu}^{-1} + \mathbf{X}^{\top} \mathbf{X})^{-1}, \end{aligned} \tag{3.7}$$

sendo $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^{\top}$ a matriz de delineamento. Além disso, a distribuição condi-

cional completa de $\mathbf{z}$ é normal truncada n -variada, com os componentes, z_i , $i = 1, \dots, n$, variáveis aleatórias normais truncadas e independentes

$$z_i | y_i, \boldsymbol{\beta}, \lambda, w_i \propto \begin{cases} \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i, 1) I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i, 1) I(z_i \leq 0) & \text{caso contrário.} \end{cases} \quad (3.8)$$

A distribuição condicional completa de $\mathbf{w}$ é normal n -variada truncada, sendo todos os seus componentes, w_i , $i = 1, \dots, n$, normais truncadas independentes a direita do zero

$$w_i | z_i, \boldsymbol{\beta}, \lambda \propto \mathcal{N}\left(-\frac{\lambda}{1 + \lambda^2}(z_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \frac{1}{1 + \lambda^2}\right) I(w_i > 0). \quad (3.9)$$

Por último, a distribuição condicional completa de λ é dada por

$$\begin{aligned} \lambda | \mathbf{z}, \boldsymbol{\beta}, \mathbf{w} &\sim \mathcal{N}(m, \nu) \\ m &= \nu \left[\tau^{-1} \alpha - \mathbf{w}^\top (\mathbf{z} - \mathbf{X} \boldsymbol{\beta}) \right] \\ \nu &= (\tau^{-1} + \mathbf{w}^\top \mathbf{w})^{-1}. \end{aligned} \quad (3.10)$$

Como nos modelos probito e logístico, existe uma forte correlação *a posteriori* entre $\boldsymbol{\beta}$ e $\mathbf{z}$. Além disso, há também uma forte correlação entre λ e $\mathbf{w}$, claramente indicadas em (3.6). Na próxima seção vamos tentar diminuir estas correlações nas cadeias fazendo o uso de atualizações conjuntas (em blocos) de algumas quantidades desconhecidas.

3.2 Algoritmos de simulação

A escolha de blocos para implementação do algoritmo Amostrador de Gibbs pode ser feita de várias maneiras no modelo probito-assimétrico. Em particular, temos o caso onde todos os blocos são escalares. O uso de gerações multivariadas em um bloco pode fornecer melhoramentos na velocidade da convergência da cadeia quando agrupa-se variáveis altamente correlacionadas. Isso ocorre porque o bloco incorpora a estrutura de correlação entre seus componentes.

Não existe uma regra geral para a escolha da formação ótima de blocos. Entretanto, blocos para os quais é fácil amostrar das distribuições condicionais completas, formam blocos naturais. Alguns desses esquemas são apresentados nas próximas seções.

3.2.1 Atualização conjunta de $\{\mathbf{z}, \boldsymbol{\beta}\}$

O método de atualização conjunta das variáveis auxiliares $\mathbf{z}$ e dos coeficientes regressores $\boldsymbol{\beta}$ para o modelo probito-assimétrico é uma extensão do que foi apresentado no capítulo anterior para os modelos probito e logito. Neste caso, atualizamos $\{\mathbf{z}, \boldsymbol{\beta}\}$ conjuntamente dado $\{\mathbf{y}, \mathbf{w}, \boldsymbol{\delta}\}$ utilizando a seguinte fatoração

$$\pi(\boldsymbol{\beta}, \mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda) = \pi(\mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda) \pi(\boldsymbol{\beta} | \mathbf{z}, \mathbf{w}, \lambda).$$

A distribuição $\pi(\boldsymbol{\beta} | \mathbf{z}, \mathbf{w}, \lambda)$ é dada em (3.7), mas agora é necessário obter a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \mathbf{w}, \lambda\}$. Mostramos no Apêndice C que $(\mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda)$ tem a seguinte densidade de probabilidade

$$\pi(\mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda) = C \phi_n(\mathbf{z}; \mathbf{X}\mathbf{b} - \lambda\mathbf{w}, I_n + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top), \quad \mathbf{z} \in R(\mathbf{y}, \mathbf{z}), \quad (3.11)$$

sendo

$$\begin{aligned} C^{-1} &= \bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \mathbf{X}\mathbf{b} - \lambda\mathbf{w}, I_n + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top) \text{ e} \\ R(\mathbf{y}, \mathbf{z}) &= \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}, \end{aligned}$$

onde $\phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a fdp de uma distribuição normal n -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, e $\bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) = IP[\mathbf{z} \in R(\mathbf{y}, \mathbf{z})]$ é a sua função de distribuição acumulada na região $R(\mathbf{y}, \mathbf{z})$. Note que $R(\mathbf{y}, \mathbf{z})$ é a mesma região apresentada nos modelos probito e logito.

Como foi dito anteriormente, é difícil simular de uma distribuição normal multivariada truncada. Assim, obtivemos as distribuições condicionais completas e utilizamos um outro algoritmo de Gibbs. As distribuições condicionais são dadas a seguir (provas no Apêndice C),

$$z_i | \mathbf{z}_{-i}, y_i, \mathbf{w}, \lambda \propto \begin{cases} \mathcal{N}(m_i, \nu_i) I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(m_i, \nu_i) I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.12)$$

onde $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. O parâmetro de localização m_i e o de escala ν_i , $i = 1, \dots, n$, são dados por

$$m_i = \mathbf{x}_i^\top \mathbf{b} - \lambda w_i + \frac{1}{1 - h_{ii}} \sum_{\substack{k=1 \\ k \neq i}}^n h_{ik} (z_k - \mathbf{x}_k^\top \mathbf{b} + \lambda w_k) \text{ e } \nu_i = \frac{1}{1 - h_{ii}},$$

onde h_{ik} denota o i -ésimo elemento da k -ésima coluna da matriz $\mathbf{H} = \mathbf{X} \mathbf{V} \mathbf{X}^\top$, com $\mathbf{V}$ definido em (3.7). Uma alternativa eficiente de obter o parâmetro de localização m_i utilizando operações matriciais é dada por

$$m_i = \mathbf{x}_i^\top \mathbf{B} - \lambda w_i - \frac{h_i}{1 - h_i} \left[z_i - \left(\mathbf{x}_i^\top \mathbf{B} - \lambda w_i \right) \right],$$

sendo z_i o valor atual da i -ésima observação do vetor $\mathbf{z}$, h_i denota o i -ésimo elemento da diagonal da matriz $\mathbf{H}$ e $\mathbf{B}$ é dado em (3.7). Como $\mathbf{B}$ é função das variáveis auxiliares z_i , devemos recalcular $\mathbf{B}$ para cada atualização de algum z_i utilizando a relação

$$\mathbf{B} = \mathbf{B}^{ant} + \mathbf{s}_i (z_i - z_i^{ant}),$$

onde $\mathbf{B}^{ant}$ e z_i^{ant} denotam, respectivamente, os valores armazenados das atualizações anteriores de $\mathbf{B}$ e z_i , e $\mathbf{s}_i$ é o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V} \mathbf{X}^\top$.

Uma amostra *a posteriori* de $\{\boldsymbol{\beta}, \lambda, \mathbf{z}, \mathbf{w}\}$ pode ser obtida nesse esquema através das atualizações sucessivas:

- i) gera-se de $(\boldsymbol{\beta}, \mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda)$ usando as distribuições (3.7) e (3.12), em seguida;
- ii) gera-se de $(\mathbf{w} | \mathbf{z}, \boldsymbol{\beta}, \lambda)$ usando (3.9), e por fim;
- iii) gera-se de $(\lambda | \mathbf{z}, \boldsymbol{\beta}, \mathbf{w})$ usando (3.10) e retorna a (i).

Este algoritmo será denotado por $HH(\mathbf{z}, \boldsymbol{\beta})$ e seu pseudo-código é apresentado no Apêndice A.

3.2.2 Atualização conjunta de $\{\mathbf{z}, \mathbf{w}\}$

A atualização conjunta do bloco de variáveis auxiliares $\{\mathbf{z}, \mathbf{w}\}$ é realizada no modelo probito-assimétrico de modo similar ao Algoritmo II no modelo logístico. Neste caso, a distribuição *a posteriori* de $\{\mathbf{z}, \mathbf{w}\}$ dado $\{\boldsymbol{\beta}, \lambda\}$ pode ser fatorada da seguinte forma

$$\pi(\mathbf{z}, \mathbf{w} | \mathbf{y}, \boldsymbol{\beta}, \lambda) = \pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \lambda) \pi(\mathbf{w} | \mathbf{z}, \boldsymbol{\beta}, \lambda),$$

onde $\pi(\mathbf{w} | \mathbf{z}, \boldsymbol{\beta}, \lambda)$ é dada em (3.9). No Apêndice C, mostramos que a distribuição $\pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \lambda)$ é normal-assimétrica n -variada truncada, onde cada componente z_i , $i =$

$1, \dots, n$, tem distribuição normal-assimétrica truncada independentemente dos demais. Mais precisamente,

$$z_i | y_i, \boldsymbol{\beta}, \lambda \propto \begin{cases} \mathcal{SN}(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda) I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{SN}(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda) I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.13)$$

sendo $\mathcal{SN}(\mu, \sigma^2, \lambda)$ a função densidade da distribuição normal-assimétrica com parâmetro de localização μ , de escala σ^2 e de forma λ (ver Apêndice D). Podemos amostrar desta distribuição normal-assimétrica truncada univariada usando o algoritmo descrito em Devroye (1986) para amostrar de distribuições truncadas. Para isso, utilizaremos apenas a fda da distribuição normal-assimétrica e sua inversa. Estas funções estão implementadas no pacote “sn” (Azzalini, 2006) disponível no programa estatístico R Development Core Team (2006).

Este algoritmo será denotado por $HH(\mathbf{z}, \mathbf{w})$ e seu pseudo-código é apresentado no Apêndice A.

3.2.3 Atualização conjunta de $\{\mathbf{z}, \lambda\}$

Outra opção de atualização conjunta para o modelo probito-assimétrico é considerar o bloco $\{\mathbf{z}, \lambda\}$. Neste caso, a distribuição conjunta *a posteriori* de $\{\mathbf{z}, \lambda\}$ condicionada à $\{\boldsymbol{\beta}, \mathbf{w}\}$ pode ser fatorada como

$$\pi(\mathbf{z}, \lambda | \mathbf{y}, \boldsymbol{\beta}, \mathbf{w}) = \pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \mathbf{w}) \pi(\lambda | \mathbf{z}, \boldsymbol{\beta}, \mathbf{w}),$$

onde $\pi(\lambda | \mathbf{z}, \boldsymbol{\beta}, \mathbf{w})$ já foi apresentada em (3.10). Mostramos no Apêndice C que $\pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \mathbf{w})$ é uma distribuição normal n -variada truncada dada por

$$\mathbf{z} | \mathbf{y}, \mathbf{w}, \boldsymbol{\beta} \propto \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w}, \mathbf{I}_n + \tau^{-1}\mathbf{w}\mathbf{w}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (3.14)$$

sendo $\text{Ind}(\mathbf{y}, \mathbf{z})$ a mesma função indicadora apresentada em (2.5). Novamente utilizamos o Amostrador de Gibbs para gerar de (3.14). A distribuição condicional completa de cada componente z_i , $i = 1, \dots, n$, (prova no Apêndice C) é dada por

$$z_i | \mathbf{z}_{-i}, y_i, \mathbf{w}, \lambda \propto \begin{cases} \mathcal{N}(m_i, \nu_i) I(z_i > 0) & \text{se } y_i = 0, \\ \mathcal{N}(m_i, \nu_i) I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.15)$$

onde $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima componente removida. Os parâmetros m_i e ν_i são dados por

$$m_i = \mathbf{x}_i^\top \boldsymbol{\beta} - w_i m - \frac{h_i}{1 - h_i} \left[z_i - \left(\mathbf{x}_i^\top \boldsymbol{\beta} - w_i m \right) \right] \quad \text{e} \quad \nu_i = \frac{1}{1 - h_i},$$

sendo z_i o valor atual da i -ésima observação do vetor $\mathbf{z}$, m é dado em (3.10) e h_i é o i -ésimo elemento da diagonal da matriz $H = \nu \mathbf{w} \mathbf{w}^\top$, com ν definido em (3.10). Logo $h_i = \nu w_i^2$.

O valor m deve ser recalculado para cada atualização de algum z_i utilizando a relação $m = m^{ant} + s_i (z_i - z_i^{ant})$, sendo m^{ant} e z_i^{ant} os valores armazenados das atualizações anteriores de m e z_i , respectivamente, e $s_i = \nu w_i$.

Esse algoritmo será denotado por $HH(\mathbf{z}, \lambda)$ e seu pseudo-código é apresentado no Apêndice A.

3.2.4 Atualização conjunta de $\{\mathbf{z}, \boldsymbol{\beta}, \lambda\}$

O modelo de regressão probito-assimétrico apresentado em (3.6) pode ser representado de forma similar ao modelo probito dado em (2.6) quando consideramos que λ é um coeficiente regressor associado as variáveis auxiliares $\mathbf{w}$. Fazendo $\mathbf{a}_i^\top = [\mathbf{x}_i^\top, w_i]$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^\top$, podemos representar esse modelo da seguinte forma

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{caso contrário.} \end{cases} \\ z_i &= \mathbf{a}_i^\top \boldsymbol{\theta} + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, 1) \\ \boldsymbol{\theta} &\sim \mathcal{N}_{p+1}(\mathbf{b}, \boldsymbol{\nu}). \end{aligned} \tag{3.16}$$

Note que esse modelo é análogo a um modelo latente. Temos também que, dado $\mathbf{w} = (w_1, \dots, w_n)^\top$, o modelo (3.16) é similar ao modelo probito definido em (2.7) com λ fazendo o papel de um coeficiente regressor.

Para a atualização de $\{\mathbf{z}, \boldsymbol{\theta}\}$, utilizamos a fatoração

$$\pi(\mathbf{z}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{w}) = \pi(\mathbf{z} | \mathbf{y}, \mathbf{w}) \pi(\boldsymbol{\theta} | \mathbf{z}, \mathbf{w}).$$

A distribuição de $\boldsymbol{\theta}$ condicionada à $\{\mathbf{z}, \mathbf{w}\}$ ainda é normal e dada por

$$\begin{aligned}\boldsymbol{\theta}|\mathbf{z}, \mathbf{w} &\sim \mathcal{N}_{p+1}(\mathbf{B}, \mathbf{V}), \text{ com} \\ \mathbf{B} &= \mathbf{V} \left(\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{A}^\top \mathbf{z} \right) \\ \mathbf{V} &= (\boldsymbol{\nu}^{-1} + \mathbf{A}^\top \mathbf{A})^{-1},\end{aligned}\tag{3.17}$$

com $\mathbf{A} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top, \dots, \mathbf{a}_n^\top)$. A distribuição $\pi(\mathbf{z}|\mathbf{y}, \mathbf{w})$ é normal multivariada truncada dada por (ver Apêndice C)

$$\mathbf{z}|\mathbf{y}, \mathbf{w} \propto \mathcal{N}_n(\mathbf{X}\mathbf{b}, \mathbf{I}_n + \mathbf{A}\boldsymbol{\nu}\mathbf{A}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}).\tag{3.18}$$

Podemos amostrar da distribuição (3.18) usando novamente um algoritmo de Gibbs, através das seguintes distribuições condicionais (ver Apêndice C)

$$z_i|\mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda \propto \begin{cases} \mathcal{N}(m_i, \nu_i)I(z_i > 0) & \text{se } y_i = 0, \\ \mathcal{N}(m_i, \nu_i)I(z_i \leq 0) & \text{caso contrário,} \end{cases}\tag{3.19}$$

onde $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. Os parâmetros de localização m_i e escala ν_i são dados por

$$m_i = \mathbf{a}_i^\top \mathbf{B} - \frac{h_i}{1 - h_i} (z_i - \mathbf{x}_i^\top \mathbf{B}) \quad \text{e} \quad \nu_i = \frac{1}{1 - h_i},$$

onde $\mathbf{B}$ é definida em (3.17), z_i denota o valor atual de z_i , h_i é o i -ésimo elemento da diagonal da matriz $\mathbf{H} = \mathbf{A}\mathbf{V}\mathbf{A}^\top$, sendo $\mathbf{V}$ definida em (3.17).

A atualização de $\mathbf{B}$ é realizada a cada atualização de algum z_i usando a relação

$$\mathbf{B} = \mathbf{B}^{ant} + \mathbf{s}_i (z_i - z_i^{ant}),$$

sendo $\mathbf{B}^{ant}$ e z_i^{ant} , os valores armazenados das atualizações anteriores de $\mathbf{B}$ e z_i , respectivamente, e $\mathbf{s}_i$ denota o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V}\mathbf{A}^\top$.

A matriz $\mathbf{A} = [\mathbf{X}, \mathbf{w}]$ é função das variáveis auxiliares $\mathbf{w}$. Assim, devemos atualizá-la a cada atualização deste vetor de variáveis. A atualização de $\mathbf{w}$ é feita através da distribuição $\pi(\mathbf{w}|\mathbf{z}, \boldsymbol{\beta}, \lambda)$ apresentada em (3.9).

Esse algoritmo será denotado por $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ e seu pseudo-código é apresentado no Apêndice A.

3.3 Comparação dos algoritmos

Neste capítulo apresentamos cinco algoritmos:

- i) AC : uso das condicionais completas de todos os parâmetros e das variáveis auxiliares;
- ii) $HH(\mathbf{z}, \boldsymbol{\beta})$: atualização conjunta das variáveis auxiliares $\mathbf{z}$ e dos coeficientes regressores $\boldsymbol{\beta}$;
- iii) $HH(\mathbf{z}, \mathbf{w})$: atualização conjunta das variáveis auxiliares $\mathbf{z}$ e $\mathbf{w}$;
- iv) $HH(\mathbf{z}, \lambda)$: atualização conjunta das variáveis auxiliares $\mathbf{z}$ e do parâmetro de assimetria λ ;
- v) $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$: atualização conjunta das variáveis auxiliares $\mathbf{z}$ e do vetor completo de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^\top$.

Nas seções 3.3.2 e 3.3.3 apresentamos, respectivamente, os resultados da análise de eficiência destes algoritmos considerando λ conhecido e desconhecido. Utilizando algumas medidas de eficiência definidas em 3.3.1 analisamos a eficiência dos algoritmos para os dados da mortalidade de insetos apresentados em Bliss (1935). Este conjunto de dados refere-se ao número de insetos adultos mortos após 5 horas de exposição ao gás carbônico para várias concentrações deste gás. Estes dados foram estudados em vários trabalhos que propuseram ligações assimétricas em regressão binária, por exemplo, Prentice (1976), Stukel (1988) e Bazán (2005). Todos estes trabalhos concluíram que modelos assimétricos são mais convenientes que modelos simétricos para o ajuste desses dados.

3.3.1 Medidas de eficiência

Utilizamos duas medidas de eficiência para ilustrar o ganho no uso dos algoritmos de atualizações conjunta em relação ao algoritmo AC . Estas medidas também foram utilizadas por Holmes e Held (2006) para mostrar que os algoritmos de atualizações conjunta para os modelos probito e logístico apresentados no Capítulo 2 são mais eficientes que o esquema tradicional (algoritmo AC). A primeira medida é a distância média Euclidiana entre atualizações do vetor de parâmetros nas iterações, definida como

$$Dist = \frac{1}{M-1} \sum_{i=1}^{M-1} \|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i+1)}\|, \quad (3.20)$$

onde $\|\cdot\|$ denota a norma Euclidiana e $\theta^{(i)}$ denota o i -ésimo vetor de uma amostra de tamanho M da distribuição *a posteriori* de θ obtida pelo método MCMC. Essa distância informa como a cadeia está se misturando. Valores grandes de $Dist$ indicam uma maior mistura na cadeia.

A segunda medida é o tamanho efetivo da amostra (*effective sample size* : ESS) descrito em Kass *et al* (1998), para maiores detalhes veja Neal (1993). Essa medida é definida por

$$ESS = \frac{M}{1 + 2 \sum_{s=1}^{\infty} \rho(s)},$$

sendo $\rho(s)$ a s -ésima autocorrelação serial. O ESS pode ser interpretado como o número de observações de uma amostra aleatória simples que estima um parâmetro de interesse com a mesma precisão de uma amostra correlacionada de tamanho M obtida via MCMC. A seguir apresentamos uma motivação para esta interpretação.

Seja $\theta_1, \dots, \theta_{ESS}$ uma amostra aleatória de tamanho ESS de uma variável aleatória θ . A esperança e a variância da média amostral $\bar{\theta} = \frac{1}{ESS} \sum_{i=1}^{ESS} \theta_i$ são dadas por

$$\mathbb{E}(\bar{\theta}) = \mathbb{E}(\theta) \quad \text{e} \quad \text{Var}(\bar{\theta}) = \frac{1}{ESS} \text{Var}(\theta). \quad (3.21)$$

Considere também que $\theta_1^*, \dots, \theta_M^*$ é uma amostra de tamanho M de θ , obtida pelo algoritmo de Gibbs. Essas observações geradas pelo Algoritmo de Gibbs são correlacionadas e a esperança e a variância de $\bar{\theta}^* = \frac{1}{M} \sum_{i=1}^M \theta_i^*$, são dadas por

$$\mathbb{E}(\bar{\theta}^*) = \mathbb{E}(\theta) \quad \text{e} \quad \text{Var}(\bar{\theta}^*) = \left[1 + 2 \sum_{s=1}^{M-1} \left(1 - \frac{s}{M} \right) \rho(s) \right] \frac{1}{M} \text{Var}(\theta). \quad (3.22)$$

Quando usamos MCMC, principalmente o Amostrador de Gibbs, as autocorrelações $\rho(s), s = 1, \dots$, são tipicamente positivas, deste modo, $\sum_{s=1}^{M-1} (1 - \frac{s}{M}) \rho(s) \geq 0$, e conseqüentemente, $\text{Var}(\bar{\theta}) \leq \text{Var}(\bar{\theta}^*)$.

Para M grande ($M \rightarrow \infty$), a variância em (3.22) pode ser reescrita como

$$\text{Var}(\bar{\theta}^*) = \frac{\tau}{M} \text{Var}(\theta), \quad (3.23)$$

onde $\tau = 1 + 2 \sum_{s=1}^{\infty} \rho(s)$ denota o tempo de autocorrelação.

As estimativas da média de θ terão a mesma precisão para as duas amostras quando

as variâncias das médias amostrais forem iguais. Portanto, igualando as variâncias das equações (3.21) e (3.23), temos que

$$ESS = \frac{M}{\tau}, \text{ sendo } \tau = 1 + 2 \sum_{s=1}^{\infty} \rho(s). \quad (3.24)$$

A estimação do ESS usando autocorrelações amostrais pode ser problemática porque a estimação de τ é prejudicada quando s cresce. Bazán (2005) e Holmes e Held (2006) assumem que a partir de um lag $k + 1$, as autocorrelações são desprezíveis e estimam o ESS baseados somente nas k primeiras autocorrelações. Uma alternativa ao uso da função de autocorrelação é utilizar a função espectral, definida em séries temporais como

$$f(\lambda) = \text{Var}(\theta) \sum_{s=-\infty}^{\infty} \rho(s) \cos(s\lambda) = \text{Var}(\theta) \left[1 + 2 \sum_{s=1}^{\infty} \rho(s) \cos(s\lambda) \right].$$

Note que a função espectral $f(\lambda)$ avaliada no ponto $\lambda = 0$ e dividida pela $\text{Var}(\theta)$ é igual ao tempo de autocorrelação τ , como definido em (3.23). Portanto, o ESS também pode ser calculado da seguinte forma,

$$ESS = \frac{M}{\tau} = \frac{M}{f(0)} \text{Var}(\theta).$$

Uma estimativa para a função espectral pode ser obtida ajustando um modelo autoregressivo de alguma ordem p à cadeia $\theta_1^*, \dots, \theta_M^*$ e, em seguida, estimando a sua respectiva função espectral no ponto zero. A variância $\text{Var}(\theta)$ é obtida de acordo com o modelo autoregressivo adotado. É claro que a precisão para a estimativa do ESS vai depender da adequabilidade do modelo autoregressivo ajustado. Este é o método utilizado para o cálculo do ESS no pacote “coda” (Plummer *et al*, 2006) e neste trabalho.

3.3.2 Análise de eficiência, com λ conhecido

O parâmetro de assimetria pode ser considerado fixo (conhecido) quando temos o conhecimento dele de estudos anteriores ou quando desejamos estimá-lo utilizando uma grade. Para o uso do grade, considera-se vários valores fixos para λ e ajusta-se o modelo, após o ajuste, verifica-se qual modelo foi o mais adequado utilizando alguma medida de ajuste, como por exemplo, Critério de Informação do Desvio (DIC), Critério de Informação Esperado de Akaike ($EAIC$), soma de quadrados dos resíduos, dentre outros. No caso de λ fixo, podemos utilizar apenas os algoritmos AC , $HH(\mathbf{z}, \boldsymbol{\beta})$ e $HH(\mathbf{z}, \mathbf{w})$ para obter uma

amostra *a posteriori* dos parâmetros regressores β . Note que $HH(z, \beta)$ e $HH(z, \beta, \lambda)$ são equivalentes quando assumimos λ fixo.

Para cada um dos três algoritmos, foram consideradas oito amostras simuladas de tamanho 20000 com período de aquecimento (*Burn-In*) de 20000 iterações. O preditor linear do modelo é definido por $\eta_i = \beta_0^* + \beta_1(x_i - \bar{x})$, com $\beta_0 = \beta_0^* - \beta_1\bar{x}$, sendo x_i a dose recebida pelo i -ésimo inseto e $\bar{x}$ a média das doses. Foi considerada uma distribuição normal vaga *a priori* para o vetor de parâmetros, isto é, $\beta^* = (\beta_0^*, \beta_1)^T \sim \mathcal{N}_2(\mathbf{0}, 1000\mathbf{I}_2)$. Para o parâmetro de assimetria, consideramos $\lambda = 4$, pois é próximo da mediana da distribuição *a posteriori* deste parâmetro obtida com os algoritmos apresentados nesta dissertação quando consideramos ele livre.

Para cada uma das cadeias, monitoramos os gráficos das médias ergódicas e aplicamos os testes de Gelman-Rubin e de Geweke nas 20000 iterações após o período de aquecimento para verificar a convergência. O teste de Gelman-Rubin estima o fator de redução de escala (R) baseado em análise de variâncias, e o teste de Geweke divide a cadeia em três partes e faz um teste de igualdade de médias entre as partes extremas usando técnicas de séries temporais. Valores de R próximos de um sugerem que a cadeia atingiu a convergência. Estes métodos são descritos com detalhes em Paulino, Turkman e Murteira (2003). Neste estudo, o maior valor obtido para a estatística R foi 1,005 e o teste de Geweke não rejeitou a hipótese de igualdade das médias com uma confiança de 95% para todas as cadeias. Esses valores indicam que os valores simulados podem ser considerados amostras das distribuições *a posteriori*.

Os programas foram escritos em linguagem de programação S e implementados no programa estatístico R Development Core Team (2006). A escolha desse programa foi motivada por ser de código aberto e por ter implementado vários pacotes úteis para o nosso trabalho, por exemplo, o pacote “sn” (Azzalini, 2006) para geração de amostras de uma variável normal-assimétrica e para os cálculos das funções de distribuições acumulada e acumulada inversa da distribuição normal-assimétrica. Além desse, o pacote “coda” (Plummer *et al*, 2006) é bastante útil para a verificação das convergências das cadeias.

A Tabela 3.1 apresenta os resultados das performances dos três algoritmos em relação às medidas ESS e $Dist$. A segunda coluna mostra o tempo de máquina (CPU), em segundos, para geração de uma amostra de tamanho 1000. A terceira e a quarta mostram as médias

dos valores de ESS e $Dist$ obtidos nas oito cadeias. As duas últimas mostram a eficiência relativa da nova abordagem de atualização conjunta em relação ao algoritmo AC . Os valores dos desvios-padrão do $Dist$ foram todos menores que 0,01 e os desvios-padrão do ESS foram aproximadamente 5% das suas médias para todos os algoritmos.

Algoritmo	CPU(s)	Médio		Relativo	
		ESS	$Dist$	ESS	$Dist$
AC	38	422,87	1,68	1,00	1,00
$HH(\mathbf{z}, \boldsymbol{\beta})$	41	945,39	2,67	2,26	1,55
$HH(\mathbf{z}, \mathbf{w})$	45	395,94	1,68	0,94	1,00

Tabela 3.1 Valores do tempo de sistema (CPU), em segundos, e das medidas ESS e $Dist$ para os diferentes algoritmos

Pela Tabela 3.1, notamos que os algoritmos $HH(\mathbf{z}, \mathbf{w})$ e AC são praticamente similares em relação as medidas utilizadas, no entanto, o algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$ é muito mais eficiente que ambos. O algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$ tem um ganho de eficiência maior que 120% segundo à medida ESS e um ganho maior que 50% em relação à medida $Dist$. Esses resultados mostram que as correlações diminuíram, com isso as cadeias obtidas pelo algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$ misturaram-se mais rapidamente e não precisamos de um tamanho de amostra tão grande quanto o necessário para o algoritmo AC obter uma boa precisão das estimativas. Assim, recomendamos o uso do algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$ quando o parâmetro de assimetria é fixo.

3.3.3 Análise de eficiência, com λ desconhecido

No caso do parâmetro de assimetria livre, podemos utilizar qualquer um dos cinco algoritmos apresentados neste capítulo. Para isso, devemos adotar um modelo probabilístico para os parâmetros. Consideramos neste estudo que o parâmetro de assimetria tem uma distribuição normal vaga *a priori*. O modelo probabilístico para o parâmetro de assimetria será dado por $\lambda \sim \mathcal{N}(0, 1000)$ e as distribuições *a priori* dos coeficientes regressores neste modelo são as mesma apresentadas para o caso de λ fixo. Entretanto, no Algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$, assumimos que $\boldsymbol{\theta} = (\beta_0, \beta_1, \lambda)^\top \sim \mathcal{N}_3(\mathbf{0}, 1000\mathbf{I}_3)$.

Para cada um dos algoritmos foram consideradas oito amostras simuladas de tamanho

100000. O maior período de aquecimento foi de 500000 iterações. O monitoramento da convergência foi feita através dos gráficos de médias ergódicas e dos testes de Gelman-Rubin e de Geweke. Esses métodos não rejeitaram a hipótese de convergência da cadeia após o período de aquecimento. Além disso, a análise desse mesmo conjunto de dados feita por Bazán *et al* (2005) usando o algoritmo *AC* no programa WinBugs forneceu resultados bem próximos aos obtidos para cada uma das 40 cadeias simuladas nesse estudo.

Os resultados das performances dos cinco algoritmos em relação às medidas *Dist* e *ESS* são apresentados nas Tabelas 3.2 e 3.3, respectivamente. A Tabela 3.2 refere-se a média e a eficiência relativa da medida *Dist* para todos os algoritmos apresentados. A Tabela 3.3 apresenta o tempo de máquina (CPU), em segundos, para geração de uma amostra de tamanho 1000 na segunda coluna, nas próximas três colunas, apresenta as médias do *ESS* para cada tipo de parâmetro do modelo, e nas duas últimas, mostra a média geral e a eficiência relativa da medida *ESS* para a abordagem de atualização conjunta.

Algoritmo	<i>Dist</i>	<i>Dist</i> Relativo
<i>AC</i>	1,68	1,00
$HH(\mathbf{z}, \boldsymbol{\beta})$	2,24	1,33
$HH(\mathbf{z}, \mathbf{w})$	1,68	1,00
$HH(\mathbf{z}, \lambda)$	1,68	1,00
$HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$	2,63	1,56

Tabela 3.2 Valores de *Dist* para os diferentes algoritmos

Considerando os resultados das Tabelas 3.2 e 3.3, observamos que houve ganho com o emprego da nova abordagem de atualização conjunta, principalmente com o uso do Algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$. Podemos ver pela Tabela 3.2 que as cadeias misturaram-se mais rapidamente nos algoritmos $HH(\mathbf{z}, \boldsymbol{\beta})$ e $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$. A Tabela 3.3 mostra ganho relativo no tamanho efetivo da amostra (*ESS*) em todos os algoritmos de atualização conjunta. No algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ o ganho é maior que 160%. Isto era esperado, pois atualizamos conjuntamente o maior número de variáveis. Nós recomendamos o uso do algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ para obtenção de uma amostra das distribuições *a posteriori* quando o parâmetro de assimetria é desconhecido.

Algoritmo	CPU(s)	<i>ESS</i>				
		β_0	β_1	λ	Médio	Relativo
<i>AC</i>	64	47,53	49,23	34,16	43,64	1,00
<i>HH(z, β)</i>	69	63,58	66,04	42,77	57,46	1,32
<i>HH(z, w)</i>	72	60,13	62,54	46,19	56,28	1,29
<i>HH(z, λ)</i>	92	79,13	82,60	40,58	67,44	1,54
<i>HH(z, β, λ)</i>	87	118,89	119,14	104,98	114,34	2,62

Tabela 3.3 Valores do tempo de sistema (CPU), em segundos, e ESS para os diferentes algoritmos

Capítulo 4

Análise de resíduos

As técnicas de diagnóstico são ferramentas indispensáveis para verificação da adequação do modelo ajustado, em particular, os resíduos são usados para verificação de possíveis afastamentos das suposições feitas para o modelo. Além disso, os resíduos também são úteis para indicar a presença de observações discrepantes, que podem provocar interferência desproporcional nos resultados inferenciais. Na abordagem clássica de modelos de regressão binária, a detecção de pontos discrepantes é usualmente baseada no resíduo ordinário $y_i - \hat{p}_i$, sendo $\hat{p}_i = F(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ a i -ésima observação ajustada e $\hat{\boldsymbol{\beta}}$ o estimador de máxima verossimilhança de $\boldsymbol{\beta}$. Existem diversos tipos de resíduos na literatura, dentre esses, os resíduos de Pearson, componente de desvio e o de Anscombe (vide, por exemplo, McCullagh e Nelder 1989). Entretanto, estes resíduos não têm distribuições amostrais conhecidas devido a natureza discreta da variável resposta. Isso prejudica a interpretação dos gráficos de resíduos e a detecção de observações discrepantes. Na abordagem Bayesiana, Albert e Chib (1995) propuseram um tipo de resíduo em regressão binária denominado resíduo latente, que está intimamente ligado ao uso de variáveis auxiliares apresentado nos capítulos anteriores.

Neste capítulo descrevemos alguns tipos de resíduos utilizados em modelos de regressão binária bayesiana. Dentre esses, apresentamos resíduos latentes para modelos obtidos através de mistura no parâmetro de escala de normais. Para o modelo probito-assimétrico, propomos o uso de dois tipos de resíduos latentes. O primeiro generaliza o resíduo latente proposto por Albert e Chib (1995) para modelos simétricos, e o segundo é baseado na relação entre as distribuições normal-assimétrica e qui-quadrado (ver Apêndice D).

4.1 Resíduos binários bayesianos

Os resíduos em modelos de regressão baseiam-se principalmente na comparação entre a i -ésima variável resposta e sua esperança. Em regressão binária, a variável resposta y_i tem

esperança $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta})$. Essa esperança é uma probabilidade não observável, pois é função do vetor de parâmetros desconhecidos $\boldsymbol{\beta}$.

O primeiro tipo de resíduo binário aqui considerado será o resíduo $r_i = y_i - p_i$. Este resíduo tem função de distribuição contínua e é função do vetor de parâmetros $\boldsymbol{\beta}$, logo, o conhecimento *a posteriori* que temos sobre $\boldsymbol{\beta}$ será refletido na distribuição *a posteriori* dos resíduos. O suporte do resíduo r_i é o intervalo $(y_i - 1, y_i)$. Assim, se o valor de y_i e a distribuição *a posteriori* de p_i estão em conflito, a distribuição *a posteriori* do resíduo estará concentrada nos valores extremos. A observação $y_i = 0$ será considerada discrepante se a distribuição *a posteriori* de r_i estiver concentrada próxima do valor -1 . Por outro lado, a observação $y_i = 1$ será considerada discrepante se a distribuição *a posteriori* de r_i estiver concentrada próxima do valor 1 .

Nos capítulos 2 e 3 foram descritas algumas formas de obtenção de uma amostra da distribuição *a posteriori* do vetor de parâmetros desconhecidos $\boldsymbol{\theta}$ para os modelos probito e logístico ($\boldsymbol{\theta} = \boldsymbol{\beta}$), e para o modelo probito-assimétrico [$\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^\top$]. Seja $\{\boldsymbol{\theta}^{(t)}\}_{t=1}^T$ uma amostra de tamanho T da distribuição *a posteriori* de $\boldsymbol{\theta}$. Como y_i e x_i são conhecidos, para obter uma amostra destes resíduos, basta considerar $r_i^{(t)} = y_i - F_{\lambda^{(t)}}(\mathbf{x}_i^\top \boldsymbol{\beta}^{(t)})$, $t \leq T$. A partir dessa amostra podemos obter medidas descritivas e densidades estimadas das distribuições *a posteriori* dos resíduos. Albert e Chib (1995) propõem o uso de gráficos do tipo boxplot para representar as amostras geradas das distribuições *a posteriori* dos resíduos contra as probabilidades ajustadas. Tais gráficos considerados conjuntamente ajudam a visualizar os resíduos discrepantes.

4.2 Resíduos latentes

Uma maneira alternativa de definir resíduos bayesianos é baseada no uso de variáveis auxiliares. Nos capítulos anteriores foram apresentadas representações do modelo de regressão binária dado em (1.1) utilizando variáveis auxiliares. O modelo (3.3) apresenta uma representação geral para modelos de regressão binária que abrange todos os modelos simétricos e assimétricos apresentados nos capítulos 2 e 3.

Podemos definir vários tipos de resíduos latentes sob o modelo (3.3). O primeiro resíduo

latente que será considerado aqui é dado por

$$\epsilon_i^*(z_i, \boldsymbol{\beta}) = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}. \quad (4.1)$$

Esse resíduo foi definido por Albert e Chib (1993) para detecção de observações discrepantes em modelos com função de ligação simétrica. Entretanto, ele também pode ser útil para o diagnóstico em modelos assimétricos.

Para modelos simétricos, o resíduo dado em (4.1) é distribuído *a priori* com a mesma distribuição utilizada para definir a função de ligação. Por exemplo, se o modelo adotado é o logito, então o resíduo (4.1) tem distribuição *a priori* logística padrão. Por outro lado, se o modelo é assimétrico, o resíduo (4.1) também tem distribuição *a priori* assimétrica. Entretanto, observamos que o parâmetro de assimetria desta distribuição tem sinal contrário ao da distribuição utilizada para definir a função de ligação. Por exemplo, no modelo probito-assimétrico dado em (3.6), $F(u) = \Phi_{CDS}(u, \lambda)$ e o resíduo $\epsilon_i^*(z_i, \boldsymbol{\beta})$ tem fda $\Phi_{CDS}(\cdot, -\lambda)$, seguindo a distribuição da variável latente z_i . Note que a distribuição do resíduo ϵ_i^* é função do parâmetro de assimetria λ . Se λ não for conhecido, a distribuição desse resíduo será função da distribuição *a priori* de λ . Por exemplo, quando λ é normalmente distribuído, a distribuição *a priori* de ϵ_i^* é desconhecida e a interpretação do tamanho dos resíduos é prejudicada. Uma alternativa é utilizar uma estimativa pontual para esse parâmetro ou definir um resíduo que não dependa de parâmetros *a priori*. Deste modo, definimos o seguinte resíduo

$$\epsilon_i(z_i, w_i, \boldsymbol{\beta}, \lambda) = z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i). \quad (4.2)$$

Este resíduo tem distribuição simétrica F com suporte nos reais. Se o modelo adotado é simétrico, o parâmetro de assimetria λ é igual a zero e o resíduo dado em (4.2) se reduz ao resíduo definido em (4.1). Se o modelo adotado é o probito-assimétrico, o resíduo ϵ_i tem distribuição normal padrão.

Para entender melhor como as observações y_i alteram a distribuição *a posteriori* dos resíduos, considere a distribuição *a posteriori* do resíduo (4.2) condicionado à $\{\boldsymbol{\beta}, \lambda\}$. A densidade desse resíduo é dada por

$$f(\epsilon_i | \mathbf{y}, \boldsymbol{\beta}, \lambda) = \begin{cases} \frac{f(\epsilon_i)}{F(\mathbf{x}_i^\top \boldsymbol{\beta})} I(\epsilon_i > -\mathbf{x}_i^\top \boldsymbol{\beta}) & \text{se } y_i = 1, \\ \frac{f(\epsilon_i)}{F(-\mathbf{x}_i^\top \boldsymbol{\beta})} I(\epsilon_i < -\mathbf{x}_i^\top \boldsymbol{\beta}) & \text{se } y_i = 0, \end{cases} \quad (4.3)$$

sendo f a função densidade associada a fda F definida em (4.2). A distribuição *a posteriori* apresentada acima tem distribuição F truncada, onde o ponto de truncamento é o negativo do preditor linear $\mathbf{x}_i^\top \boldsymbol{\beta}$. A distribuição *a posteriori* do resíduo latente ϵ_i será significamente diferente da distribuição *a priori* somente se as observações y_i estão com sinais contrários ao do preditor linear. A probabilidade desse resíduo ser maior que um valor pré-especificado $k > |-\mathbf{x}_i^\top \boldsymbol{\beta}|$ é dada por

$$\mathbb{P}(|\epsilon_i| > k | \mathbf{y}, \boldsymbol{\beta}, \lambda) = \begin{cases} \frac{1 - F(k)}{F(\mathbf{x}_i^\top \boldsymbol{\beta})} I(\epsilon_i > -\mathbf{x}_i^\top \boldsymbol{\beta}) & \text{se } y_i = 1, \\ \frac{F(-k)}{F(-\mathbf{x}_i^\top \boldsymbol{\beta})} I(\epsilon_i < -\mathbf{x}_i^\top \boldsymbol{\beta}) & \text{se } y_i = 0. \end{cases} \quad (4.4)$$

Por exemplo, no modelo logístico, assumindo que $y_i = 1$, a probabilidade desse resíduo exceder k é dada por

$$\mathbb{P}(\epsilon_i > k | y_i = 1, \boldsymbol{\beta}) = \frac{1 + e^{-\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^k}.$$

Mas, a probabilidade *a priori* deste evento ocorrer é igual a $1/(1 + e^k)$. Portanto, estas probabilidades estarão mais próximas quanto mais próximo de zero for $e^{-\mathbf{x}_i^\top \boldsymbol{\beta}}$, isto é, quanto maior for o preditor linear $\mathbf{x}_i^\top \boldsymbol{\beta}$.

No modelo probito, quando $y_i = 1$, a esperança e a variância *a posteriori* de ϵ_i são dadas por

$$\mathbb{E}(\epsilon_i | y_i = 1, \boldsymbol{\beta}) = d_i \text{ e } \text{Var}(\epsilon_i | y_i = 1, \boldsymbol{\beta}) = 1 - d_i(\mathbf{x}_i^\top \boldsymbol{\beta} + d_i),$$

respectivamente, sendo $d_i = \phi(\mathbf{x}_i^\top \boldsymbol{\beta}) / \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})$. Estes momentos serão significamente diferentes dos momentos *a priori*, $\mathbb{E}(\epsilon_i) = 0$ e $\text{Var}(\epsilon_i) = 1$, quando d_i for grande, ou equivalentemente quando o preditor linear $\mathbf{x}_i^\top \boldsymbol{\beta}$ é menor que alguma constante negativa C .

Nos modelos simétricos em geral, os resíduos $\epsilon_i(z_i, \boldsymbol{\beta})$, $i = 1, \dots, n$, podem ter variâncias *a priori* diferentes de um. Portanto, para modelos de mistura no parâmetro de escala de normais, ou seja, quando $F \in \mathcal{F}_S$, com $\mathcal{F}_S$ definido em (2.1), definimos o seguinte resíduo

$$\tau_i(\boldsymbol{\beta}, \delta_i) = \frac{z_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sqrt{\delta_i(\psi)}}, \quad (4.5)$$

sendo $\delta(\psi)$ e ψ dados em (2.4). O resíduo (4.5) condicionado à $\{\boldsymbol{\beta}, \delta_i\}$ tem distribuição normal padrão. Portanto, podemos utilizar esse resíduo e assumir que um ponto é discrepante com o mesmo critério utilizado para o caso probito, isto é, assumiremos que uma observação será discrepante se a probabilidade desse resíduo em módulo ultrapassar 1,96 for alta.

Este valor foi escolhido devido a normalidade da distribuição *a priori*. Esta probabilidade está em torno de 0,05.

Considerando agora que a função de ligação é assimétrica, isto é, $F_\lambda \in \mathcal{F}_\mathcal{A}$, onde $\mathcal{F}_\mathcal{A}$ é definida em (3.1), o resíduo ϵ_i é função de dois conjuntos de variáveis auxiliares e do vetor de parâmetros desconhecidos. Isso pode provocar uma variabilidade maior no resíduo que a existente no modelo. Para tentar reduzir essa variabilidade, propomos o uso do seguinte resíduo

$$\zeta_i(z_i, \boldsymbol{\beta}, \lambda) = \frac{(z_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{1 + \lambda^2}. \quad (4.6)$$

Considerando $F_\lambda(\cdot) = \Phi_{CDS}(\cdot, \lambda)$, obtemos o modelo probito-assimétrico dada em (3.6). Neste caso, temos que $z_i \sim \mathcal{SN}(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda)$ (veja Apêndice D). Portanto, usando as propriedades das distribuição normal-assimétrica apresentadas no Apêndice D, o resíduo $\zeta_i(z_i, \boldsymbol{\beta}, \lambda)$ condicionado a $\{\boldsymbol{\beta}, \lambda\}$ tem distribuição qui-quadrado com um grau de liberdade.

Amostras dos resíduos latentes apresentados neste capítulo podem ser facilmente obtidas para os modelos probito, logito e probito-assimétrico, utilizando-se os algoritmos apresentados nos capítulos 2 e 3. No Apêndice A, apresentamos os pseudo-códigos para gerar estas amostras.

4.3 Aplicação

Para ilustrar o comportamento dos resíduos apresentados, trabalhamos com um conjunto de dados simulados a partir de um modelo que gerasse algumas observações discrepantes. Utilizamos um modelo com probabilidade de sucesso dada por $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta})$, sendo F a fda de uma distribuição t -assimétrica, cuja função densidade é dada por

$$f(z) = \frac{2}{\sqrt{1 + \lambda^2}} t_v \left(\frac{z}{\sqrt{1 + \lambda^2}} \right) T_{\nu+1} \left[\lambda z \sqrt{\frac{\nu + 1}{(1 + \lambda^2)\nu + z^2}} \right], \quad (4.7)$$

onde t_α e T_α são, respectivamente, as fda e fdp de uma distribuição t -Student padrão com α graus de liberdade. Esta distribuição foi estudada no caso multivariado por Azzalini e Capitanio (2003).

A amostra foi obtida utilizando-se as seguinte etapas:

- Passo 1: Escolhe-se os valores do vetor de parâmetros $\boldsymbol{\theta}$. No nosso caso $\boldsymbol{\theta} = (\boldsymbol{\beta}, \nu, \lambda)^\top$,

sendo $\beta = (\beta_0, \beta_1)^\top$ os parâmetros regressores, ν os graus de liberdade e λ o parâmetro de assimetria;

- Passo 2: Seleciona-se as probabilidades de sucesso $\mathbf{p} = (p_1, \dots, p_n)^\top$;
- Passo 3: Calcula-se $\mathbf{x}_i = (1, x_i)^\top, i = 1, \dots, n$, para a função de ligação F^{-1} do seguinte modo

$$\mathbf{x}_i^\top \beta = F^{-1}(p_i) \Rightarrow x_i = [F^{-1}(p_i) - \beta_0] / \beta_1;$$

- Passo 4: Gera-se $y_i \sim \text{Bernoulli}(p_i)$, onde $p_i, i = 1, \dots, n$, são as probabilidades selecionadas no passo 2.

No passo 2, pode-se selecionar as probabilidades de sucesso de várias formas. Chen (2004) sugere escolher os p_i 's igualmente espaçados. Neste trabalho, para que os dados simulados estejam mais próximos da realidade, decidiu-se não selecionar os p_i 's igualmente espaçados. Assim como tipicamente se observa em modelos de dose-resposta, desejamos que as covariáveis $x_1, \dots, x_n$ sejam crescentes e igualmente espaçadas. Para isso, procedemos da seguinte maneira: (i) escolhemos a menor e a maior probabilidade de sucesso, denotadas aqui por p_1 e p_n , respectivamente; (ii) calculamos os valores iniciais $x_1^0 = F^{-1}(p_1)$ e $x_n^0 = F^{-1}(p_n)$; (iii) calculamos os valores iniciais restantes da seguinte forma, $x_i^0 = x_{i-1}^0 + (x_n^0 - x_1^0)/(n - 1)$, $i = 2, \dots, n - 1$, e por fim; (iv) calculamos as probabilidades de sucesso $p_i = F(\beta_0 + x_i^0 \beta_1)$, $i = 1, \dots, n$.

O conjunto de dados simulados contém 30 observações geradas a partir da distribuição t -assimétrica com os seguintes valores para os parâmetros $\beta = (0, 20)^\top$, $\nu = 5$ e $\lambda = 2$. A menor e a maior probabilidade de sucesso escolhida para o modelo foram $p_1 = 0,05$ e $p_{30} = 0,95$. O conjunto de dados gerados é apresentado na Tabela 4.1.

Ajustamos os modelos probito, logístico e probito-assimétrico para o conjunto de dados apresentados na Tabela 4.1. Foi considerada a seguinte distribuição *a priori* para β , $\mathcal{N}_2(\mathbf{0}, 1000\mathbf{I}_n)$. Para o modelo probito-assimétrico, fixamos $\lambda = 2$. Nós utilizamos os algoritmos de atualização conjunta descritos nos capítulos anteriores e simulamos uma amostra de tamanho 2000 para cada modelo. O período de aquecimento para todos os algoritmos foram de 30000 iterações. Para reduzir a autocorrelação dos valores simulados, saltamos de 10 em 10 iterações no modelo probito e de 15 em 15 nos modelos logístico e probito-assimétrico. As estatísticas são apresentados na Tabela 4.2.

Índice	x_i	y_i	Índice	x_i	y_i	Índice	x_i	y_i
1	-0,041	0	11	0,072	0	21	0,185	1
2	-0,030	0	12	0,083	1	22	0,196	0
3	-0,019	0	13	0,095	0	23	0,208	1
4	-0,007	1	14	0,106	1	24	0,249	1
5	0,004	1	15	0,117	0	25	0,230	1
6	0,015	0	16	0,128	1	26	0,242	1
7	0,027	0	17	0,140	1	27	0,253	1
8	0,038	0	18	0,151	1	28	0,264	1
9	0,049	0	19	0,162	0	29	0,276	1
10	0,061	0	20	0,174	1	30	0,287	1

Tabela 4.1 *Conjunto de dados simulados*

Modelo	Parâmetros	Média	Mediana	95% HPD	
				Inferior	Superior
Probit	β_0	-0,96	-0,95	-1,82	-0,16
	β_1	10,09	9,90	3,96	16,37
Logito	β_0	-1,69	-1,64	-3,30	-0,30
	β_1	17,37	16,94	6,10	28,26
Probit-Assimétrico	β_0	0,18	0,18	-1,01	1,35
	β_1	14,85	14,68	5,56	24,16

Tabela 4.2 *Estatísticas a posteriori dos parâmetros regressores para os modelos ajustados*

Uma análise de resíduos é realizada para tentar detectar possíveis observações discrepantes em cada modelo ajustado. As Figuras 4.1, 4.2 e 4.3 apresentam, respectivamente, as distribuições marginais *a posteriori* dos resíduos $r_i = y_i - p_i$ para os modelos probito, logito e probito-assimétrico. Estas figuras mostram as distribuições dos 30 resíduos, utilizando boxplots paralelos das distribuições dos resíduos contra as probabilidades ajustadas. Nós ajustamos as probabilidades através das médias *a posteriori* das probabilidades de sucesso.

A representação gráfica através de boxplot é útil para informar sobre a variabilidade e a simetria da distribuição. A caixa dos boxplots apresentados correspondem aos quartis, e

os valores extremos das linhas tracejadas correspondem ao 5° e ao 95° percentis. Note que as distribuições dos resíduos apresentados nas Figuras 4.1, 4.2 e 4.3 não cruzam a linha pontilhada correspondente ao valor zero. Isto ocorre porque os suportes dos resíduos r_i estão determinados pelos valores das variáveis resposta. O suporte de r_i é o intervalo $(0, 1)$ se $y_i = 1$, ou é o intervalo $(-1, 0)$, se $y_i = 0$.

Uma observação é suspeita de ser discrepante se a distribuição do seu correspondente resíduo estar localmente distante do zero. Uma forma de verificar isso, é calcular a probabilidade desse resíduo ser maior em módulo que alguma constante positiva k . Albert e Chib (1995) sugerem usar $k = 0,75$. Adicionamos linhas pontilhadas paralelas nos valores $-0,75$ e $0,75$ nas Figuras 4.1, 4.2 e 4.3 para visualizar a probabilidade desses resíduos exceder estes valores. Observe que, se o boxplot de um resíduo em particular não ultrapassar essas linhas, então a probabilidade dessa observação ser discrepante é menor que 0,05. Observações com alta probabilidade de serem discrepantes correspondem à resíduos com boxplots que cruzam significamente essas linhas. As Figuras 4.1, 4.2 e 4.3, e a quarta coluna das Tabelas 4.3, 4.4 e 4.5, mostram que existem quatro observações com probabilidades altas de serem discrepantes. As mesmas observações foram detectadas nos três modelos ajustados e correspondem as observações 3 e 4 ($y_3 = y_4 = 1$), e as observações 19 e 22 ($y_{19} = y_{22} = 0$).

Figura 4.1 *Boxplots das distribuições a posteriori dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo probito*

Figura 4.2 *Boxplots das distribuições a posteriori dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo logito*

Figura 4.3 *Boxplots das distribuições a posteriori dos resíduos $r_i = y_i - p_i$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo probito-assimétrico*

Embora o resíduo r_i seja útil para detecção de possíveis observações discrepantes, é difícil comparar as distribuições de dois resíduos com diferentes dispersões (amplitudes), ou sob modelos diferentes. A distribuição *a priori* do resíduo r_i não é conhecida devido a natureza discreta da variável resposta. Assim, um problema é definir qual o melhor valor para o ponto de corte k .

Os resíduos latentes descritos neste capítulo são apresentados nas Figuras 4.4, 4.5, 4.6, 4.7, 4.8 e 4.9. Estas figuras mostram os boxplots paralelos das distribuições *a posteriori* dos resíduos latentes contra as probabilidades ajustadas. As Figuras 4.4, 4.5 e 4.6 mostram, respectivamente, os boxplots da distribuição *a posteriori* do resíduo latente $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ para os modelos probito, logito e probito-assimétrico. Os resíduos ϵ_i^* 's para os modelos probito e logito são simétricos e distribuídos *a priori* com distribuições normal padrão e logística padrão, respectivamente. Uma observação y_i poderá ser considerada discrepante nestes modelos se a distribuição *a posteriori* de seu respectivo resíduo ϵ_i^* está localmente distante do zero. Alternativamente, outra forma de verificar se uma observação y_i é discrepante, é calcular a probabilidade do evento $\{|\epsilon_i^*| > k\}$, onde k é tal que a probabilidade *a priori* desse evento ocorrer é baixa, em torno de 0,05. O valor de k tal que $\mathbb{P}(|\epsilon_i^*| > k | \mathbf{y})$ é aproximadamente 0,05 é dado por $k = \Phi^{-1}(0,975) = 1,96$ no modelo probito e $k = 3,66$ no modelo logito (pois as caudas da distribuição logística são mais pesadas que a da distribuição normal). Adicionamos linhas paralelas nos pontos k e $-k$ das Figuras 4.4 e 4.5 para visualizar melhor essa probabilidade. As probabilidades *a posteriori* de um resíduo, individualmente, ultrapassar esses valores no modelos probito e logito são apresentadas na quinta coluna das Tabelas 4.3 e 4.4, respectivamente.

Em nosso estudo, o parâmetro de assimetria do modelo probito-assimétrico é fixo igual a dois, o que implica que o resíduo $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ tem distribuição normal-assimétrica *a priori*, mas o parâmetro de assimetria tem sinal contrário ao da função de ligação, isto é, $\lambda = -2$. Na Figura 4.7 mostramos os boxplots das distribuições destes resíduos contra as probabilidades ajustadas. Uma observação y_i é uma possível observação discrepante se a distribuição *a posteriori* do seu respectivo resíduo estar localmente distante da mediana da distribuição *a priori*. Como no modelo probito e logito, podemos escolher pontos de cortes $k_1 < k_2$ tais que a probabilidade de um resíduo ϵ_i^* estar entre esses pontos seja aproximadamente 0,95. Há várias maneiras de escolher tais pontos, nós escolhemos os

Figura 4.4 *Boxplots das distribuições a posteriori dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o ajuste do modelo probito*

Figura 4.5 *Boxplots das distribuições a posteriori dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo logito*

pontos que consideram caudas iguais, $k = -5$ e $k = 1, 1$. Outra alternativa seria escolher os pontos que definem o intervalo de maior densidade (HPD). Na Figura 4.6, adicionamos linhas paralelas correspondentes aos pontos $k_1 = -5$ e $k = 1, 1$. A probabilidade de um resíduo ultrapassar esses pontos é apresentada na quinta coluna da Tabela 4.5.

As observações com maiores probabilidades de serem discrepantes considerando os resíduos ϵ_i^* 's foram as mesmas obtidas na análise do resíduo anterior para todos os modelos, ou seja, as observações 4,5,19 e 22.

Figura 4.6 *Boxplots das distribuições a posteriori dos resíduos latentes $\epsilon_i^* = z_i - \mathbf{x}_i^\top \boldsymbol{\beta}$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo probito-assimétrico*

O resíduo ϵ_i^* é função do parâmetro de assimetria no modelo probito-assimétrico. Isso prejudica as análises quando esse parâmetro é desconhecido, pois neste caso, o resíduo ϵ_i^* não tem distribuição *a priori* conhecida. Uma alternativa é o uso do resíduo $\epsilon_i = z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i)$. Este resíduo não depende de parâmetros *a priori* e é normalmente distribuído no modelo probito-assimétrico. Portanto, detectamos observações discrepantes utilizando este resíduo da mesma maneira que detectamos com o resíduo ϵ_i^* no caso probito. Outro resíduo também normalmente distribuído *a priori* é o resíduo $\tau_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta})/\sqrt{\delta_i}$ no modelo logito, pois o modelo logístico pode ser obtido através de mistura no parâmetro de escala de normais. As Figuras 4.7 e 4.8 apresentam, respectivamente, os boxplots das

distribuições *a posteriori* dos resíduos ϵ_i e τ_i para os modelos probito-assimétrico e logito. A sexta coluna das Tabelas 4.4 e 4.5 mostram as probabilidades destes resíduos *a posteriori* em modulo serem maiores que 1,96.

Figura 4.7 *Boxplots das distribuições a posteriori dos resíduos latentes $\epsilon_i = z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i)$ contra as probabilidades ajustadas $\mathbb{E}(p_i|\mathbf{y})$ para o modelo probito-assimétrico*

Na Figura 4.9, apresentamos o resíduo $\zeta_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 / (1 + \lambda^2)$, este resíduo também não depende de parâmetros *a priori* no modelo probito-assimétrico, e é distribuído *a priori* com distribuição qui-quadrado com um grau de liberdade. Uma observação será considerada discrepante se seu respectivo resíduo tiver uma probabilidade alta de exceder o valor 3,84. Este valor foi escolhido pois a probabilidade de um resíduo, individualmente, exceder esse valor *a priori* está em torno de 0,05. Adicionamos uma linha pontilhada no ponto $k = 3,84$ para visualizar melhor essa probabilidade. Ela também é mostrada na sétima coluna da Tabela 4.5. As observações com probabilidade alta de serem discrepantes utilizando estes resíduos foram correspondente a fracassos; observações 19 e 22. Além dessas, as observações 13 e 15 também tiveram probabilidades relativamente altas de serem discrepantes. Estas observações estão na cauda mais leve da distribuição normal-assimétrica. Entretanto, este tipo de resíduo não interpretou as observações 4 e 5 como discrepantes.

Figura 4.8 *Boxplots das distribuições a posteriori dos resíduos latentes $\tau_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta}) / \sqrt{\delta_i}$ contra as probabilidades ajustadas $\mathbb{E}(p_i | \mathbf{y})$ para o modelo logito*

Figura 4.9 *Boxplots das distribuições a posteriori dos resíduos latentes $\zeta_i = (z_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 / (1 + \lambda^2)$ contra as probabilidades ajustadas $\mathbb{E}(p_i | \mathbf{y})$ para o modelo probito-assimétrico*

Tabela 4.3 *Observações e probabilidades de pontos discrepantes no modelo probito*

Índice	y_i	$E(p_i \mathbf{y})$	$P(r_i > 0,75 \mathbf{y})$	$P(\epsilon_i^* > 1,96 \mathbf{y})$
1	0	0,113	0,000	0,034
2	0	0,130	0,000	0,027
3	0	0,150	0,000	0,028
4	1	0,173	0,780	0,252
5	1	0,199	0,716	0,198
6	0	0,227	0,000	0,034
7	0	0,259	0,000	0,030
8	0	0,293	0,000	0,040
9	0	0,330	0,000	0,040
10	0	0,370	0,000	0,045
11	0	0,411	0,000	0,048
12	1	0,454	0,030	0,058
13	0	0,498	0,003	0,054
14	1	0,542	0,002	0,054
15	0	0,585	0,053	0,052
16	1	0,627	0,000	0,043
17	1	0,667	0,000	0,041
18	1	0,705	0,000	0,044
19	0	0,741	0,480	0,113
20	1	0,773	0,000	0,030
21	1	0,803	0,000	0,035
22	0	0,829	0,803	0,218
23	1	0,852	0,000	0,027
24	1	0,873	0,000	0,037
25	1	0,891	0,000	0,035
26	1	0,906	0,000	0,030
27	1	0,920	0,000	0,032
28	1	0,931	0,000	0,031
29	1	0,941	0,000	0,032
30	1	0,950	0,000	0,033

Tabela 4.4 *Observações e probabilidades de pontos discrepantes no modelo logito*

Índice	y_i	$E(p_i \mathbf{y})$	$P(r_i > 0,75 \mathbf{y})$	$P(\epsilon_i^* > 3,66 \mathbf{y})$	$P(\tau_i > 1,96 \mathbf{y})$
1	0	0,111	0,000	0,026	0,033
2	0	0,126	0,000	0,030	0,034
3	0	0,144	0,000	0,036	0,048
4	1	0,165	0,813	0,242	0,220
5	1	0,188	0,751	0,186	0,180
6	0	0,215	0,000	0,032	0,028
7	0	0,245	0,000	0,034	0,040
8	0	0,278	0,000	0,041	0,044
9	0	0,314	0,000	0,034	0,030
10	0	0,354	0,000	0,040	0,043
11	0	0,397	0,000	0,042	0,046
12	1	0,442	0,046	0,065	0,061
13	0	0,488	0,006	0,054	0,057
14	1	0,534	0,002	0,042	0,045
15	0	0,580	0,061	0,060	0,060
16	1	0,625	0,000	0,046	0,046
17	1	0,667	0,000	0,040	0,049
18	1	0,706	0,000	0,032	0,032
19	0	0,741	0,499	0,136	0,142
20	1	0,774	0,000	0,036	0,038
21	1	0,802	0,000	0,024	0,032
22	0	0,828	0,806	0,210	0,204
23	1	0,850	0,000	0,028	0,032
24	1	0,870	0,000	0,030	0,040
25	1	0,887	0,000	0,031	0,038
26	1	0,901	0,000	0,032	0,036
27	1	0,914	0,000	0,028	0,032
28	1	0,925	0,000	0,032	0,036
29	1	0,934	0,000	0,034	0,038
30	1	0,943	0,000	0,036	0,043

Tabela 4.5 *Observações e probabilidades de pontos discrepantes no modelo probito-assimétrico*

Índice	y_i	$E(p_i \mathbf{y})$	$P(r_i > 0,75 \mathbf{y})$	$P(\epsilon_i^* \notin (-5; 1, 1) \mathbf{y})$	$P(\epsilon_i > 1,96 \mathbf{y})$	$P(\zeta_i > 3,89 \mathbf{y})$
1	0	0,114	0,000	0,034	0,038	0,055
2	0	0,133	0,000	0,038	0,041	0,070
3	0	0,156	0,000	0,034	0,044	0,059
4	1	0,182	0,746	0,236	0,146	0,000
5	1	0,211	0,673	0,182	0,124	0,000
6	0	0,243	0,000	0,034	0,038	0,074
7	0	0,279	0,000	0,040	0,038	0,085
8	0	0,317	0,000	0,040	0,050	0,076
9	0	0,357	0,000	0,042	0,042	0,083
10	0	0,399	0,001	0,044	0,055	0,089
11	0	0,442	0,002	0,050	0,048	0,090
12	1	0,484	0,010	0,059	0,049	0,000
13	0	0,527	0,007	0,056	0,062	0,127
14	1	0,568	0,000	0,048	0,038	0,000
15	0	0,607	0,070	0,063	0,068	0,142
16	1	0,644	0,000	0,044	0,041	0,000
17	1	0,679	0,000	0,036	0,036	0,000
18	1	0,712	0,000	0,038	0,036	0,000
19	0	0,742	0,500	0,112	0,098	0,222
20	1	0,769	0,000	0,026	0,032	0,000
21	1	0,794	0,000	0,032	0,038	0,000
22	0	0,817	0,781	0,183	0,116	0,374
23	1	0,837	0,000	0,032	0,045	0,001
24	1	0,855	0,000	0,034	0,046	0,002
25	1	0,872	0,000	0,031	0,044	0,002
26	1	0,886	0,000	0,028	0,040	0,006
27	1	0,899	0,000	0,026	0,038	0,005
28	1	0,911	0,000	0,031	0,038	0,006
29	1	0,921	0,000	0,035	0,042	0,016
30	1	0,930	0,000	0,038	0,040	0,014

Capítulo 5

Considerações finais

Neste trabalho descrevemos o uso de variáveis auxiliares em modelos de regressão binária bayesiana. A introdução dessas variáveis tem como objetivo a obtenção de formas conhecidas para as distribuições condicionais completas, as quais possibilitam a implementação do algoritmo Amostrador de Gibbs. Além disso, baseado nestas variáveis auxiliares, podemos definir diversos tipos de resíduos que são úteis para detectar a presença de observações discrepantes. No entanto, a introdução dessas variáveis pode produzir cadeias de Monte Carlo fortemente correlacionadas, fato este que prejudica a convergência do algoritmo. Nós apresentamos propostas de agrupamento das quantidades desconhecidas em blocos viáveis para a simulação conjunta, obtendo algoritmos mais eficientes. Para os modelos probito e logístico, apresentamos e implementamos as propostas de atualização conjunta dadas por Holmes e Held (2006). Para a ligação probito-assimétrico propomos quatro diferentes maneiras de construir os blocos e comparamos estes algoritmos através de duas medidas de eficiência, a distância média Euclidiana entre atualizações e o tamanho efetivo da amostra. Concluímos que os novos algoritmos propostos forneceram cadeias menos correlacionadas e que se misturam mais rapidamente que as cadeias obtidas através do algoritmo AC (sem blocos). Em particular, o algoritmo que mostrou a melhor performance no modelo probito-assimétrico com o parâmetro de assimetria fixo foi o $HH(\mathbf{z}, \boldsymbol{\beta})$, ele proporcionou ganho superior a 120% no tamanho efetivo da amostra. Enquanto que, para o parâmetro de assimetria livre, os algoritmos $HH(\mathbf{z}, \boldsymbol{\beta})$ e $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ foram os mais eficientes, sendo que o $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ forneceu um aumento em torno de 160% no tamanho efetivo da amostra em relação ao AC . Esses resultados mostram que as correlações diminuíram, e com isso, as cadeias obtidas pelos algoritmos propostos misturam-se mais rapidamente. Mostrou também que não precisamos de um tamanho de amostra tão grande quanto o necessário para o algoritmo AC , quando desejamos obter uma boa precisão das estimativas.

Finalmente, adaptamos e implementamos os resíduos propostos para a ligação probito para os modelos logístico e probito-assimétrico. Para um conjunto de dados simulados, detectamos as mesmas observações discrepantes utilizando os resíduos r_i , ϵ_i^* , ϵ_i e τ_i para os diferentes modelos. Entretanto, o resíduo latente ζ_i (resíduo que tem distribuição *a priori* qui-quadrado no modelo probito-assimétrico) não interpretou as observações localizadas em uma das caudas da distribuição como discrepantes.

Os pseudo-códigos para todos os algoritmos descritos estão disponíveis no Apêndice A, e visam facilitar a implementação computacional dos algoritmos em qualquer linguagem de programação. Eles mostram como obter amostras das distribuições *a posteriori* de todas as quantidades desconhecidas e resíduos apresentados neste trabalho. Optamos pelo uso do programa R Development Core Team (2006).

Este trabalho oferece várias possibilidades para futuras pesquisas, entre estas podemos destacar:

- Extensões para respostas multinomiais com ligação probito-assimétrico.
- Discussão do problema de seleção de variáveis em modelos com ligação probito-assimétrico.
- Extensões dos algoritmos para modelos mistos binários.
- Construção de uma biblioteca para o programa estatístico R que rode estas funções através do C.

Apêndice A

Pseudo-Códigos

Apresenta-se neste apêndice os pseudo-códigos para a implementação dos algoritmos de Gibbs em blocos dados nos capítulos anteriores. Considera-se nos pseudo-códigos que: $A[i]$ denota o i -ésimo elemento da matriz coluna A ; $A[i, j]$ denota o i -ésimo elemento da j -ésima linha da matriz A ; $A[i,]$ e $A[, j]$ denotam, respectivamente, a i -ésima linha e a j -ésima coluna de A ; AB denota o produto matricial entre A e B ; $A[i,]B[, j]$ denota o produto matricial entre a i -ésima linha de A pela j -ésima coluna de B . Linhas comentadas são precedidas por $##$.

A.1 Procedimento para amostragem no modelo probito

O código para o modelo probito assume *a priori* que, $\beta \sim \mathcal{N}_p(\mathbf{b}, \nu)$ e a matriz de delineamento X tem dimensão $(n \times p)$.

A.1.1 Algoritmo HH

$##$ Armazena-se as constantes inalteradas dentro do MCMC

$V \leftarrow (X^\top X + v^{-1})^{-1}$ $##$ V denota a matriz de covariâncias de β

$S \leftarrow VX^\top$

PARA $j = 1$ até número de observações

$H[j] \leftarrow X[j,]S[, j]$

$T[j] \leftarrow H[j]/(1 - H[j])$

$Q[j] \leftarrow T[j] + 1$

FIM

$##$ Inicializa-se as variáveis latentes Z , de normais truncadas independentes

$Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$

$B \leftarrow V(v^{-1}b + X^\top Z)$ $##$ B denota a média condicional de β

```

PARA  $i = 1$  até número de iterações
  PARA  $j = 1$  até número de observações
     $Z_{ant} \leftarrow Z[j]$ 
     $m \leftarrow X[j, ]B$ 
     $m \leftarrow m - T[j](Z[j] - m)$ 
    ## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
     $Z[j] \sim \mathcal{N}(m, Q[j])\text{Ind}(Y[j], Z[j])$ 
    ## Atualiza-se  $B$  através da seguinte relação
     $B \leftarrow B + (Z[j] - Z_{ant})S[, j]$ 
  FIM
  ## Amostra-se de  $\beta$  a partir da distribuição normal  $p$ -variada
   $\beta[, i] \sim \mathcal{N}_p(B, V)$ 
  ## Obtém-se os resíduos
  PARA  $j = 1$  até número de observações
     $r[i, j] \leftarrow Y[i] - \Phi(X[j, ]\beta[, i])$ 
     $\epsilon^*[i, j] \leftarrow Z[j] - X\beta[, i]$ 
  FIM
FIM das iterações MCMC

```

A.2 Procedimentos para amostragem no modelo logístico

Considera-se no código que, $\beta \sim \mathcal{N}_p(\mathbf{b}, \nu)$ *a priori* e a matriz de delineamento X tem dimensão $(n \times p)$. O pseudo-código para geração de uma amostra *a posteriori* da variável de mistura no modelo logístico também é apresentado nesta seção.

A.2.1 Algoritmo I

```

## Inicializa-se as variáveis de mistura
 $\Delta \leftarrow \mathbf{I}_n$ 
## Inicializa-se as variáveis latentes  $Z$ , de normais truncadas independentes
 $Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(Y, Z)$ 
PARA  $i = 1$  até número de iterações

```

$V \leftarrow (X^\top \Delta^{-1} X + v^{-1})^{-1}$
 $S \leftarrow V X^\top$
 $B \leftarrow V(v^{-1}b + X^\top \Delta^{-1} Z)$ ## B denota a média condicional de β
 PARA $j = 1$ até número de observações
 $Z_{ant} \leftarrow Z[j]$
 $H[j] \leftarrow X[j,] S[, j]$
 $W[j] \leftarrow H[j] / (\Delta[j, j] - H[j])$
 $m \leftarrow X[j,] B$
 $m \leftarrow m - W[j] (Z[j] - m)$
 $q \leftarrow \Delta[j, j] (W[j] + 1)$
 ## Amostra-se de $Z[j]$ a partir da distribuição normal truncada
 $Z[j] \sim \mathcal{N}(m, q) \text{Ind}(Y[j], Z[j])$
 ## Atualiza-se B através da seguinte relação
 $B \leftarrow B + \left(\frac{Z[j] - Z_{ant}}{\Delta[j, j]} \right) S[, j]$
 FIM
 ## Amostra-se de β a partir da distribuição normal p -variada
 $\beta[, i] \sim \mathcal{N}_p(B, V)$
 ## Agora, amostra-se das variáveis de mistura
 PARA $j = 1$ até número de observações
 $R \leftarrow (Z[j] - X[j,] \beta[, i])$
 $\Delta[j, j] \sim \pi(\delta | \mathbf{z}, \beta)$
 ## Ver programa A.2.3.

FIM

Obtém-se os Resíduos

PARA $j = 1$ até número de observações

$r[i, j] \leftarrow Y[i] - F_{Lo}(X[j,] \beta[, i])$ ## F_{Lo} denota a fda de uma distribuição $Lo(0, 1)$
 $\epsilon^*[i, j] \leftarrow Z[j] - X[j,] \beta[, i]$
 $\tau[i, j] \leftarrow \epsilon^*[i, j] / \sqrt{\Delta[j, j]}$

FIM

FIM das iterações MCMC

A.2.2 Algoritmo II

```

## Inicializa-se as variáveis de mistura
 $\Delta \leftarrow \mathbf{I}_n$ 
## Inicializa-se as variáveis latentes  $Z$ , de logísticas truncadas independentes
 $Z \sim Lo_n(0, 1)Ind(Y, Z)$ 
PARA  $i = 1$  até número de iterações
     $V \leftarrow (X^\top \Delta^{-1} X + v^{-1})^{-1}$ 
     $B \leftarrow V(v^{-1}b + X^\top \Delta^{-1} Z)$ 
    ## Amostra-se os valores de  $\beta$  a partir da distribuição normal  $p$ -variada
     $\beta[, i] \sim \mathcal{N}_p(B, V)$ 
    ## Agora, atualiza-se conjuntamente  $\{Z, \Delta\}$ 
    PARA  $j = 1$  até número de observações
         $m \leftarrow X[j, ]\beta[, i]$ 
        ## Amostra-se de  $Z[j]$ , da distribuição logística truncada
         $Z[j] \sim Lo(m, 1)Ind(Y[j], Z[j])$ 
        ## Amostra-se das variáveis de mistura
         $R \leftarrow (Z[j] - X[j, ]\beta[, i])$ 
         $\Delta \sim \pi(\delta | \mathbf{z}, \beta)$ 
    ## Ver programa A.2.3.
FIM
## Obtém-se os resíduos
PARA  $j = 1$  até número de observações
     $r[i, j] \leftarrow Y[i] - F_{Lo}(X[j, ]\beta[, i])$  ##  $F_{Lo}$  denota a fda de uma distribuição  $Lo(0, 1)$ 
     $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\beta[, i]$ 
     $\tau[i, j] \leftarrow \epsilon^*[i, j] / \sqrt{\Delta[j, j]}$ 
FIM
FIM das iterações MCMC

```

A.2.3 Procedimento para amostrar de $\delta \sim \pi(\delta | \mathbf{z}, \beta)$

REPITA

Temos que, $r^2 = (z_i - \mathbf{x}_i^\top \beta)^2$

```

## Para iniciar, gera-se uma amostra da variável proposta
 $Y \sim \chi_1^2$ 
 $Y \leftarrow 1 + \left( Y - \sqrt{Y(4r + Y)} \right) / (2r)$ 
 $U \sim U[0, 1]$ 
SE  $U \leq 1/(1 + Y)$ 
    ENTÃO  $\delta \leftarrow r/Y$ 
    SENÃO  $\delta \leftarrow rY$ 
## Agora, gera-se de  $\delta \sim GIG(0, 5; 1; r^2)$ 
 $U \sim U[0, 1]$ 
SE  $\delta > 4/3$ 
    OK  $\leftarrow$  direita( $U, \delta$ )
SENÃO
    OK  $\leftarrow$  esquerda( $U, \delta$ )
ATÉ OK = 1

```

O procedimento acima chama duas funções, direita() e esquerda(), dependendo do valor proposto de δ . Os pseudo-códigos destas funções são apresentados a seguir:

O pseudo-código para a região direita

```

OK  $\leftarrow$  direita( $U, \delta$ )
 $Z \leftarrow 1$ 
 $X \leftarrow \exp(-0.5\delta)$ 
 $j \leftarrow a \leftarrow b \leftarrow 0$ 
REPITA
     $j \leftarrow j + 1$ 
     $Z \leftarrow Z - (j + 1)^2 X^{(j+1)^2 - 1}$ 
    SE  $Z > U$  ENTÃO
        OK  $\leftarrow 1$ 
         $a \leftarrow a + 1$ 
         $b \leftarrow b - 1$ 

```

FIM do SE

$j \leftarrow j + 1$

$Z \leftarrow Z + (j + 1)^2 X^{(j+1)^2-1}$

SE $Z \leq U$ ENTÃO

OK $\leftarrow 0$

$a \leftarrow a - 1$

$b \leftarrow b + 1$

FIM do SE

ATÉ $a = 2$ ou $b = 2$

RETORNE **OK**.

O pseudo-código para a região esquerda

OK $\leftarrow$ esquerda(U, δ)

$H \leftarrow 0.5 \log(2) + 2,5 \log(\pi) - 2,5 \log(\delta) - \frac{\pi}{2\delta} + 0,5\delta$

$lU \leftarrow \log(U)$

$Z \leftarrow 1$

$K \leftarrow \delta/\pi$

$X \leftarrow \exp(-\pi^2/(2\delta))$

$j \leftarrow a \leftarrow b \leftarrow 0$

REPITA

$j \leftarrow j + 1$

$Z \leftarrow Z - K X^{j^2-1}$

SE $H + \log(Z) > lU$ ENTÃO

OK $\leftarrow 1$

$a \leftarrow a + 1$

$b \leftarrow b - 1$

FIM do SE

$j \leftarrow j + 1$

$Z \leftarrow Z + (j + 1)^2 X^{(j+1)^2-1}$

SE $H + \log(Z) \leq lU$ ENTÃO **OK** $\leftarrow 0$

```

OK  $\leftarrow$  0
 $a \leftarrow a - 1$ 
 $b \leftarrow b + 1$ 
FIM do SE
ATÉ  $a = 2$  ou  $b = 2$ 
RETORNE OK.

```

A.3 Procedimentos para amostragem no modelo probito-assimétrico

No Capítulo 3 apresentamos quatro algoritmos de Gibbs em blocos para o modelo probito-assimétrico. Nos algoritmos $HH(\mathbf{z}, \boldsymbol{\beta})$, $HH(\mathbf{z}, \mathbf{w})$ e $HH(\mathbf{z}, \lambda)$, o código considera que $\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{b}, \boldsymbol{\nu})$ e $\lambda \sim \mathcal{N}(\alpha, \tau)$ são independentes *a priori*. No algoritmo $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$, consideramos que $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^\top \sim \mathcal{N}_{p+1}(\mathbf{b}, \boldsymbol{\nu})$ *a priori*. Considera-se também que a matriz de delineamento X tem dimensão $(n \times p)$.

A.3.1 Algoritmo $HH(\mathbf{z}, \boldsymbol{\beta})$

```

## Armazena-se as constantes inalteradas dentro do MCMC
 $V \leftarrow (X^\top X + v^{-1})^{-1}$  ##  $V$  é a matriz de covariâncias de  $\boldsymbol{\beta}$ 
 $S \leftarrow VX^\top$ 
PARA  $j = 1$  até número de observações
     $H[j] \leftarrow X[j,]S[,j]$ 
     $T[j] \leftarrow H[j]/(1 - H[j])$ 
     $Q[j] \leftarrow T[j] + 1$ 
FIM
## Inicializa-se as variáveis latentes  $Z$  e  $W$ , de normais truncadas independentes
 $Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$ 
 $W \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) I(W > \mathbf{0})$ 
PARA  $i = 1$  até número de iterações
     $B \leftarrow V [v^{-1}b + X^\top (Z - \lambda[i]W)]$  ##  $B$  denota a média condicional de  $\boldsymbol{\beta}$ 
    PARA  $j = 1$  até número de observações

```

```

 $Z_{ant} \leftarrow Z[j]$ 
 $m \leftarrow X[j, ]B - \lambda[i]W[j]$ 
 $m \leftarrow m - T[j](Z[j] - m)$ 
## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
 $Z[j] \sim \mathcal{N}(m, Q[j])Ind(Y[j], Z[j])$ 
## Atualiza-se  $B$  através da seguinte relação
 $B \leftarrow B + (Z[j] - Z_{ant})S[, j]$ 
FIM

## Amostra-se de  $\beta$  a partir da distribuição normal  $p$ -variada
 $\beta[, i] \sim \mathcal{N}_p(B, V)$ 
 $s \leftarrow 1/(1 + \lambda[i]^2)$ 
PARA  $j = 1$  até número de observações
   $r \leftarrow s\lambda[i](Z[j] - X[j, ]\beta[, i])$ 
  ## Amostra-se de  $W$  a partir de distribuições normais truncadas
   $W[j] \sim \mathcal{N}(r, s)I(W[j] > 0)$ 
FIM

 $q \leftarrow 1/(\tau^{-1} + W^\top W)$ 
 $d \leftarrow q[\tau^{-1}\alpha - W^\top(Z - X\beta[, i])]$ 
## Amostra-se de  $\lambda$  a partir da distribuição normal
 $\lambda[i] \sim \mathcal{N}(d, q)$ 
## Obtém-se os resíduos
PARA  $j = 1$  até número de observações
   $r[i, j] \leftarrow Y[i] - \Phi_{CDS}(X[j, ]\beta[, i], \lambda[i])$ 
   $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\beta[, i]$ 
   $\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$ 
   $\zeta[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$ 
FIM

FIM das iterações MCMC

```

A.3.2 Algoritmo $HH(z, w)$

```

## Armazena-se as constantes inalteradas dentro do MCMC

```


$$V \leftarrow (X^\top X + v^{-1})^{-1}$$

$$S \leftarrow VX^\top$$

PARA $j = 1$ até número de observações

$$H[j] \leftarrow X[j,]S[,j]$$

$$T[j] \leftarrow H[j]/(1 - H[j])$$

$$Q[j] \leftarrow T[j] + 1$$

FIM

Inicializa-se as variáveis latentes Z e W , de normais truncadas independentes

$$Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$$

$$W \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) I(W > \mathbf{0})$$

PARA $i = 1$ até número de iterações

$$B \leftarrow V [v^{-1}b + X^\top (Z - \lambda[i]W)] \quad \text{## Denota a média condicional de } \beta$$

Amostra-se de β a partir da distribuição normal p -variada

$$\beta[,i] \sim \mathcal{N}_p(B, V)$$

Agora, atualiza-se conjuntamente $\{Z, W\}$

PARA $j = 1$ até número de observações

$$m \leftarrow X[j,]\beta[,i] - \lambda[i]W[j]$$

Amostra-se de $Z[j]$ através da distribuição normal-assimétrica truncada

$$Z[j] \sim \mathcal{SN}(m, 1 + \lambda[i]^2, -\lambda[i]) \text{Ind}(Y[j], Z[j])$$

FIM

$$s \leftarrow 1/(1 + \lambda[i]^2)$$

PARA $j = 1$ até número de observações

$$r \leftarrow s\lambda[i](Z[j] - X[j,]\beta[,i])$$

Amostra-se de W a partir de distribuições normais truncadas

$$W[j] \sim \mathcal{N}(r, s) I(W[j] > 0)$$

FIM

$$q \leftarrow 1/(\tau^{-1} + W^\top W)$$

$$d \leftarrow q [\tau^{-1}\alpha - W^\top (Z - X\beta[,i])]$$

Amostra-se de λ a partir da distribuição normal

$$\lambda[i] \sim \mathcal{N}(d, q)$$

Obtém-se os resíduos

PARA $j = 1$ até número de observações

$$r[i, j] \leftarrow Y[i] - \Phi_{CDS}(X[j,]\beta[, i], \lambda[i])$$

$$\epsilon^*[i, j] \leftarrow Z[j] - X[j,]\beta[, i]$$

$$\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$$

$$\zeta[i, j] \leftarrow \epsilon^*[i, j]^2 / (1 + \lambda[i]^2)$$

FIM

FIM das iterações MCMC

A.3.3 Algoritmo $HH(z, \lambda)$

Armazena-se as constantes inalteradas dentro do MCMC

$$V \leftarrow (X^\top X + v^{-1})^{-1}$$

Inicializa-se as variáveis latentes Z e W , de normais truncadas independentes

$$Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$$

$$W \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n) I(W > \mathbf{0})$$

PARA $i = 1$ até número de iterações

$$S \leftarrow \tau W$$

$$q \leftarrow 1 / (\tau^{-1} + W^\top W)$$

$$d \leftarrow q [\tau^{-1} \alpha - W^\top (Z - X\beta[, i])] \quad \text{## } d \text{ denota a média condicional de } \lambda$$

PARA $j = 1$ até número de observações

$$H[j] \leftarrow W[j]S[j]$$

$$T[j] \leftarrow H[j] / (1 - H[j])$$

$$Q[j] \leftarrow T[j] + 1$$

$$Z_{ant} \leftarrow Z[j]$$

$$m \leftarrow X[j,]\beta[, i] - W[j]m$$

$$m \leftarrow m - T[j] (Z[j] - m_i)$$

Amostra-se de $Z[j]$ a partir da distribuição normal truncada

$$Z[j] \sim \mathcal{N}(m, Q[j]) \text{Ind}(Y[j], Z[j])$$

Atualiza-se d através da seguinte relação

$$d \leftarrow d + (Z[j] - Z_{ant})S[j]$$

FIM

Amostra-se de λ a partir da distribuição normal

$$\lambda[i] \sim \mathcal{N}(d, q)$$

$$s \leftarrow 1/(1 + \lambda[i]^2)$$

PARA $j = 1$ até número de observações

$$r \leftarrow s\lambda[i](Z[j] - X[j,]\beta[,i])$$

Amostra-se de W a partir de distribuições normais truncadas

$$W[j] \sim \mathcal{N}(r, s)I(W[j] > 0)$$

FIM

$$B \leftarrow V[v^{-1}b + X^\top(Z - \lambda[i]W)] \quad \text{## } B \text{ denota a média condicional de } \beta$$

$$\beta[,i] \sim \mathcal{N}_p(B, V)$$

Obtém-se os resíduos

PARA $j = 1$ até número de observações

$$r[i, j] \leftarrow Y[i] - \Phi_{CDS}(X[j,]\beta[,i], \lambda[i])$$

$$\epsilon^*[i, j] \leftarrow Z[j] - X[j,]\beta[,i]$$

$$\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$$

$$\zeta[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$$

FIM

FIM das iterações MCMC

A.3.4 Algoritmo $HH(\mathbf{z}, \beta, \lambda)$

Inicializa-se as variáveis latentes Z e W , de normais truncadas independentes

$$Z \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)I(Y, Z)$$

$$W \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)I(W > \mathbf{0})$$

Concatena a matriz X e o vetor W

$$A \leftarrow [X: - W]$$

PARA $i = 1$ até número de iterações

$$V \leftarrow (A^\top A + v^{-1})^{-1}$$

$$S \leftarrow VA^\top$$

$$B \leftarrow V(v^{-1}b + A^\top Z) \quad \text{## } B \text{ denota a média condicional de } \theta$$

PARA $j = 1$ até número de observações

$$H[j] \leftarrow A[j,]S[,j]$$

$$T[j] \leftarrow H[j]/(1 - H[j])$$

```

 $Q[j] \leftarrow T[j] + 1$ 
 $Z_{ant} \leftarrow Z[j]$ 
 $m \leftarrow A[j, ]B$ 
 $m \leftarrow m - T[j](Z[j] - m)$ 
## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
 $Z[j] \sim \mathcal{N}(m, Q[j])\text{Ind}(Y[j], Z[j])$ 
## Atualiza-se  $B$  através da seguinte relação
 $B \leftarrow B + (Z[j] - Z_{ant})S[, j]$ 

```

FIM

Amostra-se de θ a partir da distribuição normal $(n + 1)$ -variada

$\theta[, i] \sim \mathcal{N}_{p+1}(B, V)$

$\beta \leftarrow \theta[1 : p, i]$

$\lambda \leftarrow \theta[p + 1, i]$

$s \leftarrow 1/(1 + \lambda^2)$

PARA $j = 1$ até número de observações

$r \leftarrow s\lambda(Z[j] - X[j,]\beta)$

Amostra-se de W a partir de distribuições normais truncadas

$W[j] \sim \mathcal{N}(r, s)I(W[j] > 0)$

FIM

$q \leftarrow 1/(\tau^{-1} + W^\top W)$

$d \leftarrow q [\tau^{-1}\alpha - W^\top (Z - X\beta)]$

Amostra-se de λ a partir da distribuição normal

$\lambda[i] \sim \mathcal{N}(d, q)$

Obtém-se os resíduos

PARA $j = 1$ até número de observações

$r[i, j] \leftarrow Y[i] - \Phi_{CDS}(X[j,]\beta[, i], \lambda[i])$

$\epsilon^*[i, j] \leftarrow Z[j] - X[j,]\beta[, i]$

$\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$

$\zeta[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$

FIM

FIM das iterações MCMC

Apêndice B

Método para amostrar da variável misturadora no modelo logístico

Neste apêndice descrevemos como amostrar da distribuição condicional completa do vetor de variáveis misturadoras $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_n\}$ no modelo de regressão logística apresentado na seção 2.2.

As variáveis $\delta_1, \dots, \delta_n$, condicionadas à $\{\mathbf{z}, \boldsymbol{\beta}\}$, são independentes, e cada variável δ_i tem distribuição dada por

$$\pi(\delta|\mathbf{z}, \boldsymbol{\beta}) \propto \delta^{-1/2} \sum_{n=1}^{\infty} (-1)^{n+1} \exp \left\{ -\frac{1}{2} \left(\frac{r^2}{\delta} + n^2 \delta \right) \right\}, \quad (\text{B.1})$$

sendo $r^2 = (\mathbf{z}_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$. Esta distribuição não tem forma conhecida, entretanto, podemos simular uma amostra dela de forma eficiente usando um algoritmo de rejeição.

Note que a distribuição $\pi(\delta|\mathbf{z}, \boldsymbol{\beta})$ em (B.1) tem o núcleo parecido com a densidade de uma gaussiana-inversa generalizada, denotada por *GIG* (ver Apêndice D). Homes e Held (2006) sugerem o uso da *GIG* com parâmetros: $\lambda = 0, 5$, $\psi = 1$ e $\chi = r^2$, como distribuição proposta para implementação do algoritmo de rejeição.

Para o uso do algoritmo de rejeição, devemos gerar da distribuição proposta. Atkinson (1982) mostra como simular de uma distribuição *GIG* usando distribuições exponenciais. Usaremos as propriedades da distribuição gaussiana-inversa generalizada apresentadas no Apêndice D, e amostraremos desta distribuição de forma mais eficiente através da seguinte relação

$$GIG(0, 5; 1; r^2) = rGIG(0, 5; r; r) = \frac{r}{GIG(-0, 5; r; r)} = \frac{r}{GI(1; r)}, \quad (\text{B.2})$$

onde *GI* denota a distribuição gaussiana-inversa (ver Apêndice D). Assim, usando a relação (B.2), podemos amostrar de uma distribuição *GIG* através de uma distribuição *GI*. Uma amostra da *GI* pode ser obtida facilmente usando o método da inversão com raízes múltiplas (Devroyer, 1986, p. 148). No Apêndice D é apresentado este algoritmo.

Seja $g(\cdot)$ a fdp da distribuição proposta, denotada aqui $GIG(0, 5; 1; r^2)$, sendo $r^2 = (z_i - x_i^\top \beta)^2$. No algoritmo de rejeição, gera-se um valor de $g(\cdot)$, e este valor é aceito para $\pi(\cdot | \mathbf{z}, \beta)$ com probabilidade $\alpha(\cdot)$,

$$\alpha(\delta) = \frac{\pi(\delta | \mathbf{z}, \beta)}{Mg(\delta)}, \text{ sendo } M \geq \sup_{\delta} \frac{\pi(\delta | \mathbf{z}, \beta)}{g(\delta)}. \quad (\text{B.3})$$

Como a distribuição *a posteriori* de δ é proporcional ao produto de sua distribuição *a priori* pela função de verossimilhança, temos que (B.3) pode ser reescrita da seguinte forma,

$$\alpha(\delta) = \frac{l(\delta)\pi(\delta)}{Kg(\delta)}, \quad (\text{B.4})$$

onde $l(\delta)$ denota a verossimilhança, $l(\delta) \propto \delta^{-1/2} \exp(-0, 5r^2/\delta)$, K é maior ou igual ao $\sup_{\delta} \frac{l(\delta)\pi(\delta)}{g(\delta)}$, e $\pi(\delta)$ é a distribuição *a priori* de δ ,

$$\pi(\delta) = \frac{1}{4} \delta^{-1/2} KS\left(\frac{1}{2}\delta\right), \quad (\text{B.5})$$

em que $KS(\cdot)$ denota a densidade da distribuição Kolmogorov-Smirnov assintótica dada no Apêndice D. A distribuição *a priori* (B.5) vem da transformação de variáveis $\delta_i = (2\psi_i)^2$ apresentada em (2.11), e pode ser reescrita de duas formas (ver Apêndice D). A primeira forma é dada por

$$\pi(\delta) = \sum_{n=1}^{\infty} (-1)^{n+1} n^2 \exp\left\{-\frac{n^2\delta}{2}\right\}, \quad \delta > 0, \quad (\text{B.6})$$

e a segunda por

$$\pi(\delta) = \frac{\sqrt{2\pi}}{\delta} \sum_{n=1}^{\infty} \left[\frac{(2n-1)^2}{\delta^{3/2}} - \frac{1}{\delta^{1/2}} \right] \exp\left\{-\frac{(2n-1)^2}{2\delta}\right\}, \quad \delta > 0. \quad (\text{B.7})$$

Notamos que podemos assumir $K = 1$ em (B.4). Neste caso, a probabilidade de aceitação se reduz a

$$\alpha(\delta) = \exp\left(\frac{\delta}{2}\right) \pi(\delta).$$

Esta probabilidade é a soma de uma série infinita, o que torna sua resolução direta complicada. Entretanto, as densidades (B.6) e (B.7), aqui denotadas por densidade A e densidade B, respectivamente, podem ser escritas em forma de séries alternadas (Devroye, 1986, p.161-165), definida como

$$\pi(\delta) = ch(\delta) \sum_{n=0}^{\infty} (-1)^n a_n(\delta), \quad \text{com } a_0(\delta) = 1.$$

Para melhor visualizar as funções $ch(\cdot)$ e $a_n(\cdot)$, escrevemos a densidade A como

$$\pi(\delta) = \exp\left(-\frac{\delta}{2}\right) \sum_{n=1}^{\infty} (-1)^n (n+1)^2 \exp\left\{-\frac{[(n+1)^2 - 1]\delta}{2}\right\}, \quad \delta > 0,$$

que em formato de séries alternadas é dada por

$$\begin{aligned} ch(\delta) &= \exp\left(-\frac{\delta}{2}\right), \\ a_n(\delta) &= (n+1)^2 \exp\left\{-\frac{[(n+1)^2 - 1]\delta}{2}\right\}, \quad n \geq 0. \end{aligned} \quad (\text{B.8})$$

A densidade B também pode ser reescrita de tal forma que possa ser melhor visualizada as funções $ch(\cdot)$ e $a_n(\cdot)$ dadas em (B.8). Essa representação é dada por

$$\pi(\delta) = \frac{\sqrt{2\pi}\pi^2}{\delta^{5/2}} \exp\left(-\frac{\pi^2}{2\delta}\right) \sum_{n=1}^{\infty} \left[(2n-1)^2 - \frac{\delta}{\pi^2}\right] \exp\left\{-\frac{(2n-1)^2\pi^2}{2\delta}\right\}, \quad \delta > 0,$$

e novamente em forma de séries alternadas,

$$\begin{aligned} ch^*(\delta) &= \frac{\sqrt{2\pi}\pi^2}{\delta^{5/2}} \exp\left(-\frac{\pi^2}{2\delta}\right), \\ a_n^*(\lambda) &= \begin{cases} \frac{\delta}{\pi^2} \exp\left\{-\frac{(n^2-1)\pi^2}{2\delta}\right\} & \text{se } n \text{ é ímpar e } \delta > 0, \\ (n+1) \exp\left\{-\frac{[(n+1)^2-1]\pi^2}{2\delta}\right\} & \text{se } n \text{ é par e } \delta > 0. \end{cases} \end{aligned} \quad (\text{B.9})$$

Para aplicação do método de séries alternadas, devemos verificar se as séries a_n e a_n^* são monótonas decrescentes. Na expansão em séries da densidade A, temos que

$$\log\left(\frac{a_{n-1}(\delta)}{a_n(\delta)}\right) = -2\log\left(1 + \frac{1}{n}\right) + \frac{1}{2}(2n+1)\delta \geq -\frac{2}{n} + \frac{1}{2}(2n+1)\delta \geq -2 + \frac{3}{2}\delta > 0.$$

A última desigualdade é satisfeita para todo $\delta > 4/3$, e conseqüentemente monótona nesse intervalo. Para a expansão em séries da densidade B, consideramos n par. Neste caso

$$\frac{a_n^*(\delta)}{a_{n+1}^*(\delta)} = \frac{(n+1)\pi^2}{\delta} \geq \frac{\pi^2}{\delta} > 1, \quad \forall \delta \in (0, \pi^2)$$

e

$$\log\left(\frac{a_{n-1}^*(\delta)}{a_n^*(\delta)}\right) = -2\log(n+1) - \log\left(\frac{\pi^2}{\delta}\right) + 2n\frac{\pi^2}{\delta}.$$

Notamos que a última expressão também é monótona crescente em $\delta \in (0, \pi^2)$, mas com a restrição que $n \geq 2$. Verificamos também que ela não é maior que $2n - 2\log(n+1) \geq 0$.

Logo, $\pi(\delta)$ é monótona em $(4/3, \infty)$ e em $(0, \pi^2)$. O próximo passo é particionar o espaço paramétrico de δ em duas regiões disjuntas e construir séries monótonas alternadas. O ponto de quebra desse método de mistura pode ser qualquer um no intervalo $[4/3, \pi^2]$. Nós usaremos o ponto $4/3$, pois a série obtida através da densidade B converge mais rapidamente.

A probabilidade de aceitação $\alpha(\delta)$ em (B.4) é dada por

$$\alpha(\delta) = \exp\left(\frac{\delta}{2}\right) \pi(\delta) = \begin{cases} \sum_{n=0}^{\infty} (-1)^n a_n(\delta) & \text{se } \delta > \frac{4}{3}, \\ \frac{\sqrt{2\pi}\pi^2}{\delta^{5/2}} \exp\left(\frac{\delta}{2} - \frac{\pi^2}{2\delta}\right) \sum_{n=0}^{\infty} (-1)^n a_n^*(\delta) & \text{caso contrário,} \end{cases}$$

sendo a_n e a_n^* definidos em (B.9) e (B.10), respectivamente. O pseudo-código para esse algoritmo é apresentado no Apêndice A.

A Figura B.1 apresenta a taxa de aceitação estimada para valores de r^2 numa grade que varia de 10^{-10} à 25. Para cada valor da grade, foram geradas uma amostra de tamanho 50000 e estimada a taxa de aceitação.

Figura B.1 Curva da taxa de aceitação estimada para os valores de r_i^2

Pela Figura B.1, observamos que o algoritmo é menos eficiente quando $r_i^2 \rightarrow 0$. Entretanto, a taxa de aceitação é aproximadamente 0,29 para r_i^2 pequeno, em torno de 10^{-10} . Para r_i^2 próximo de um, a taxa de aceitação é aproximadamente 0,62, e para $r_i^2 > 10$ a aceitação é praticamente um.

Apêndice C

Provas das distribuições do Capítulo 3

Para as provas das distribuições obtidas no Capítulo 3, precisamos de alguns resultados envolvendo relações matriciais que são apresentados no três lemas a seguir.

Lema C.1 *Sejam A , B , C e D , matrizes de dimensões $p \times p$, $p \times n$, $n \times n$ e $n \times p$. Se todas as inversas necessárias existem, então*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

Prova: Multiplicando o lado direito por $(A + BCD)$, temos que

$$\begin{aligned} & (A + BCD) [A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}] \\ &= I_n + B [C - (I_n + CDA^{-1}B)(C^{-1} + DA^{-1}B)^{-1}] DA^{-1} \\ &= I_n + B [C - C(C^{-1} + DA^{-1}B)(C^{-1} + DA^{-1}B)^{-1}] DA^{-1} \\ &= I_n \end{aligned}$$

■

Lema C.2 *Sejam, X e ν : matrizes de dimensões $n \times p$ e $p \times p$, respectivamente; z , w , β e b : vetores de dimensões $n \times 1$, $n \times 1$, $p \times 1$ e $p \times 1$, respectivamente, e os escalares λ , α e τ . Se todas as inversas necessárias existem, então*

i) $(z - X\beta + \lambda w)^\top (z - X\beta + \lambda w) + (\beta - b)^\top \nu^{-1}(\beta - b) = (\beta - B)^\top V^{-1}(\beta - B) + C$,
sendo

$$\begin{aligned} C &= [z - (Xb - \lambda w)]^\top (I_n + X\nu X^\top)^{-1} [z - (Xb - \lambda w)] \\ B &= V \left[\nu^{-1}b + X^\top (z + \lambda w) \right] \\ V &= \left(\nu^{-1} + X^\top X \right)^{-1} \end{aligned}$$

ii) $(\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w})^\top (\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w}) + \mathbf{w}^\top \mathbf{w} = (1 + \lambda^2) (\mathbf{w} - \mathbf{m})^\top (\mathbf{w} - \mathbf{m}) + C$, sendo

$$\begin{aligned} C &= (1 + \lambda^2)^{-1} (\mathbf{z} - \mathbf{X}\beta)^\top (\mathbf{z} - \mathbf{X}\beta) \\ \mathbf{m} &= -\frac{\lambda}{1 + \lambda^2} (\mathbf{z} - \mathbf{X}\beta) \end{aligned}$$

iii) $(\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w})^\top (\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w}) + \tau^{-1}(\lambda - \alpha)^2 = \nu^{-1}(\lambda - m)^2 + C$, sendo

$$\begin{aligned} C &= [\mathbf{z} - (\mathbf{X}\beta - \alpha\mathbf{w})]^\top (\mathbf{I}_n + \tau\mathbf{w}\mathbf{w}^\top)^{-1} [\mathbf{z} - (\mathbf{X}\beta - \alpha\mathbf{w})] \\ m &= \nu \left[\tau^{-1}\alpha - \mathbf{w}^\top (\mathbf{z} - \mathbf{X}\beta) \right] \\ \nu &= \left(\tau^{-1} + \mathbf{w}^\top \mathbf{w} \right)^{-1} \end{aligned}$$

Lema C.3 *Sejam, $\mathbf{X}$ e ν : matrizes de dimensões $n \times p$ e $p \times p$, respectivamente; $\mathbf{z}$, $\mathbf{w}$, β e $\mathbf{b}$: vetores de dimensões $n \times 1$, $n \times 1$, $p \times 1$ e $p \times 1$, respectivamente, e os escalares λ , α e τ . Se todas as inversas necessárias existem, então*

$$i) (\mathbf{z} - \boldsymbol{\mu})^\top (\mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top)^{-1} (\mathbf{z} - \boldsymbol{\mu}) = (1 - h_{ii}) \left[z_i - \left(\frac{1}{1 - h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 + C,$$

onde $C = - \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (z_k - \mu_k) \left(h_{kj} - \frac{h_{ki}h_{ij}}{1 - h_{ii}} \right) (z_j - \mu_j)$; $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, sendo cada

componente $\mu_i = \mathbf{x}_i^\top \mathbf{b} - \lambda w_i$ e $h_{ij} = \mathbf{x}_i^\top (\nu^{-1} + \mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_j$.

$$ii) (\mathbf{z} - \boldsymbol{\mu})^\top (\mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top)^{-1} (\mathbf{z} - \boldsymbol{\mu}) = (1 - h_{ii}) \left[z_i - \left(\frac{1}{1 - h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 + C,$$

onde $C = - \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (z_k - \mu_k) \left(h_{kj} - \frac{h_{ki}h_{ij}}{1 - h_{ii}} \right) (z_j - \mu_j)$; $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, sendo cada

componente $\mu_i = \mathbf{x}_i^\top \beta - \alpha w_i$ e $h_{ij} = w_i^\top (\tau^{-1} + \mathbf{w}^\top \mathbf{w})^{-1} w_j$.

As provas dos Lemas C.2 e C.3 são obtidas utilizando o Lema C.1 e operações matriciais de soma e produto.

A seguir apresentamos as provas das distribuições obtidas no Capítulo 3.

Proposição C.1 *Sob o modelo probito-assimétrico dado em (3.6), temos que as distribuições condicionais completas de $(\beta, \mathbf{z}, \mathbf{w}, \lambda | \mathbf{y})$ são dadas por (3.7), (3.8), (3.9) e (3.10), respectivamente.*

Prova:

Distribuição (3.7): Utilizando o teorema de Bayes, temos que a distribuição condicional completa de β é dada por

$$\pi(\beta | \mathbf{z}, \mathbf{w}, \lambda) = C^{-1} \pi(\mathbf{z}, \beta | \mathbf{w}, \lambda), \quad \text{sendo } C = \pi(\mathbf{z} | \mathbf{w}, \lambda). \quad (\text{C.1})$$

Lembrando que as distribuições *a priori* de β , $\mathbf{w}$ e λ são independentes sob o modelo (3.6), temos que $\pi(\mathbf{z}, \beta | \mathbf{w}, \lambda) = \pi(\mathbf{z} | \beta, \mathbf{w}, \lambda) \pi(\beta)$, onde $\pi(\mathbf{z} | \beta, \mathbf{w}, \lambda)$ pode ser escrita como

$$\pi(\mathbf{z} | \beta, \mathbf{w}, \lambda) = \phi_n(\mathbf{z}; \mathbf{X}\beta - \lambda\mathbf{w}, I_n), \quad (\text{C.2})$$

em que $\phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a fdp de uma distribuição normal n -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Assim, segue que

$$\begin{aligned} \pi(\mathbf{z}, \beta | \mathbf{w}, \lambda) &= \pi(\mathbf{z} | \beta, \mathbf{w}, \lambda) \pi(\beta) \\ &= (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\beta - \lambda\mathbf{w})]^\top [\mathbf{z} - (\mathbf{X}\beta - \lambda\mathbf{w})] \right\} \\ &\quad \times (2\pi)^{-p/2} |\boldsymbol{\nu}|^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - \mathbf{b})^\top \boldsymbol{\nu}^{-1} (\beta - \mathbf{b}) \right\} \\ &= (2\pi)^{-(n+p)/2} |\boldsymbol{\nu}|^{-1/2} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[(\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w})^\top (\mathbf{z} - \mathbf{X}\beta + \lambda\mathbf{w}) + (\beta - \mathbf{b})^\top \boldsymbol{\nu}^{-1} (\beta - \mathbf{b}) \right] \right\}. \end{aligned}$$

Usando o Lema C.2, temos que

$$\pi(\mathbf{z}, \beta | \mathbf{w}, \lambda) = K \exp \left\{ -\frac{1}{2} (\beta - \mathbf{B})^\top \mathbf{V}^{-1} (\beta - \mathbf{B}) \right\}, \quad (\text{C.3})$$

sendo

$$\begin{aligned} K &= (2\pi)^{-(n+p)/2} |\boldsymbol{\nu}|^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\mathbf{b} - \lambda\mathbf{w})]^\top (\mathbf{I}_n + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top)^{-1} [\mathbf{z} - (\mathbf{X}\mathbf{b} - \lambda\mathbf{w})] \right\} \\ \mathbf{B} &= \mathbf{V} \left[\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z} + \lambda\mathbf{w}) \right] \\ \mathbf{V} &= \left(\boldsymbol{\nu}^{-1} + \mathbf{X}^\top \mathbf{X} \right)^{-1}. \end{aligned}$$

Agora, substituindo (C.3) em (C.1), temos que

$$\pi(\boldsymbol{\beta}|\mathbf{z}, \mathbf{w}, \lambda) = C^{-1}\pi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{w}, \lambda) = C^{-1}K \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta} - \mathbf{B})^\top \mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{B}) \right\}. \quad (\text{C.4})$$

Observe que C e K , dados em (C.1) e (C.3), respectivamente, não dependem de $\boldsymbol{\beta}$, e o fator restante da distribuição (C.4) é o núcleo de uma distribuição normal p -variada com vetor de médias $\mathbf{B}$ e matriz de covariâncias $\mathbf{V}$. Portanto, a função (C.4) só será uma fdp se $C^{-1}K = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}$, o que implica que $C = \pi(\mathbf{z}|\mathbf{w}, \lambda)$ é a fdp de uma distribuição normal n -variada dada por

$$\pi(\mathbf{z}|\mathbf{w}, \lambda) = \phi_n \left(\mathbf{z}; \mathbf{X}\mathbf{b} - \lambda\mathbf{w}, \mathbf{I}_n + \mathbf{X}\boldsymbol{\nu}\mathbf{X}^\top \right), \quad (\text{C.5})$$

e a distribuição condicional completa de $\boldsymbol{\beta}$ é dada por

$$\boldsymbol{\beta}|\mathbf{z}, \mathbf{w}, \lambda \sim \mathcal{N}_p(\mathbf{B}, \mathbf{V}),$$

onde $\mathbf{B}$ e $\mathbf{V}$ são dados em (C.3).

Distribuição (3.8): Utilizando novamente o teorema de Bayes, temos que a distribuição condicional completa de $\mathbf{z}$ é dada por

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \mathbf{w}, \lambda) = C^{-1}\pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\beta}, \mathbf{w}, \lambda), \quad \text{sendo } C = \pi(\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \lambda). \quad (\text{C.6})$$

Por outro lado, $\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \lambda) = \pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{w}, \lambda)\pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\beta}, \mathbf{w}, \lambda)$, onde $\pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{w}, \lambda)$ é dada em (C.2) e

$$\pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\beta}, \mathbf{w}, \lambda) = \pi(\mathbf{y}|\mathbf{z}) = \text{Ind}(\mathbf{y}, \mathbf{z}) = \prod_{i=1}^n \text{Ind}(y_i, z_i), \quad (\text{C.7})$$

sendo $\text{Ind}(y_i, z_i) = I(z_i > 0)I(y_i = 1) + I(z_i \leq 0)I(y_i = 0)$, onde $I(A)$ denota a função indicadora no conjunto A . Logo, substituindo (C.7) e (C.2) em (C.6), temos que

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \mathbf{w}, \lambda) = C^{-1}\phi_n(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w}, \mathbf{I}_n)\text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.8})$$

Note que $C = \pi(\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \lambda)$ não depende de $\mathbf{z}$ e o fator restante da distribuição (C.8) é o núcleo de uma distribuição normal n -variada truncada com vetor de médias $\mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w}$ e matriz de covariâncias $\mathbf{I}_n$. Portanto, os componentes z_i , $i = 1 \dots, n$ são variáveis aleatórias independentes com distribuição normal truncada dada por

$$z_i|y_i, w_i, \boldsymbol{\beta}, \lambda \propto \begin{cases} \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i, 1)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda w_i, 1)I(z_i \leq 0) & \text{se } y_i = 0. \end{cases}$$

Distribuição (3.9): Utilizando o teorema de Bayes, temos que a distribuição condicional completa de $\mathbf{w}$ é dada por

$$\pi(\mathbf{w}|\mathbf{z}, \boldsymbol{\beta}, \lambda) = C^{-1} \pi(\mathbf{z}, \mathbf{w}|\boldsymbol{\beta}, \lambda), \quad \text{sendo } C = \pi(\mathbf{z}|\boldsymbol{\beta}, \lambda), \quad (\text{C.9})$$

Novamente, pela independência entre as distribuições *a priori* de $\boldsymbol{\beta}$, $\mathbf{w}$ e λ , segue que

$$\begin{aligned} \pi(\mathbf{z}, \mathbf{w}|\boldsymbol{\beta}, \lambda) &= \pi(\mathbf{z}|\mathbf{w}, \boldsymbol{\beta}, \lambda) \pi(\mathbf{w}) \\ &= (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w})]^\top [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w})] \right\} \\ &\quad \times \left(\frac{2}{\pi} \right)^{n/2} \exp \left\{ -\frac{1}{2} \mathbf{w}^\top \mathbf{w} \right\} I(\mathbf{w} > \mathbf{0}) \\ &= \pi^{-n} \exp \left\{ -\frac{1}{2} \left[(\mathbf{z} - \mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{w})^\top (\mathbf{z} - \mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{w}) + \mathbf{w}^\top \mathbf{w} \right] \right\} I(\mathbf{w} > \mathbf{0}). \end{aligned}$$

Agora, usando o Lema C.2, temos que

$$\pi(\mathbf{z}, \mathbf{w}|\boldsymbol{\beta}, \lambda) = K \exp \left\{ -\frac{1 + \lambda^2}{2} (\mathbf{w} - \mathbf{m})^\top (\mathbf{w} - \mathbf{m}) \right\} I(\mathbf{w} > \mathbf{0}), \quad (\text{C.10})$$

sendo

$$\begin{aligned} K &= \pi^{-n} \exp \left\{ -\frac{1}{2(1 + \lambda^2)} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ \mathbf{m} &= -\frac{\lambda}{1 + \lambda^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

Finalmente, substituindo (C.10) em (C.9), temos que

$$\pi(\mathbf{w}|\mathbf{z}, \boldsymbol{\beta}, \lambda) = C^{-1} K \exp \left\{ -\frac{(1 + \lambda^2)}{2} (\mathbf{w} - \mathbf{m})^\top (\mathbf{w} - \mathbf{m}) \right\} I(\mathbf{w} > \mathbf{0}). \quad (\text{C.11})$$

Observe que C e K não dependem de $\mathbf{w}$, e o fator restante da distribuição (C.11) é o núcleo de uma distribuição normal p -variada positiva. Logo,

$$\mathbf{w}|\mathbf{z}, \boldsymbol{\beta}, \lambda \propto \mathcal{N}_n \left(-\frac{\lambda}{1 + \lambda^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}), (1 + \lambda^2)^{-1} \mathbf{I}_n \right) I(\mathbf{w} > \mathbf{0}).$$

Como a matriz de covariâncias $(1 + \lambda^2)^{-1} \mathbf{I}_n$ é diagonal, os componentes w_i , $i = 1, \dots, n$, são variáveis aleatórias independentes com distribuição normal truncada dada por

$$w_i|z_i, \boldsymbol{\beta}, \lambda \propto \mathcal{N} \left(-\frac{1 + \lambda^2}{2} (z_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \frac{1}{1 + \lambda^2} \right) I(w_i > 0)$$

Distribuição (3.10): Utilizando o teorema de Bayes, temos que a distribuição condicional completa de λ é dada por

$$\pi(\lambda|\mathbf{z}, \boldsymbol{\beta}, \mathbf{w}) = C^{-1} \pi(\mathbf{z}, \lambda|\boldsymbol{\beta}, \mathbf{w}), \quad \text{sendo } C = \pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{w}). \quad (\text{C.12})$$

Pela independência entre as distribuições *a priori* de $\boldsymbol{\beta}$, $\mathbf{w}$ e λ , temos que

$$\begin{aligned} \pi(\mathbf{z}, \lambda|\boldsymbol{\beta}, \mathbf{w}) &= \pi(\mathbf{z}|\mathbf{w}, \boldsymbol{\beta}, \lambda) \pi(\lambda) \\ &= (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w})]^\top [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \lambda\mathbf{w})] \right\} \\ &\quad \times (2\pi\tau)^{-1/2} \exp \left\{ -\frac{1}{2} \frac{(\lambda - \alpha)^2}{\tau} \right\} \\ &= (2\pi)^{-(n+1)/2} \tau^{-1/2} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[(\mathbf{z} - \mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{w})^\top (\mathbf{z} - \mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{w}) + \frac{(\lambda - \alpha)^2}{\tau} \right] \right\}. \end{aligned}$$

Usando o Lema C.2, segue que

$$\pi(\mathbf{z}, \lambda|\boldsymbol{\beta}, \lambda) = K \exp \left\{ -\frac{1}{2} \frac{(\lambda - m)^2}{\tau} \right\}, \quad (\text{C.13})$$

sendo

$$\begin{aligned} K &= (2\pi)^{-(n+1)/2} \tau^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w})]^\top (\mathbf{I}_n + \tau\mathbf{w}\mathbf{w}^\top)^{-1} [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w})] \right\} \\ m &= \nu \left[\tau^{-1}\alpha - \mathbf{w}^\top (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right] \\ \nu &= \left(\mathbf{w}^\top \mathbf{w} + \tau^2 \right)^{-1} \end{aligned}$$

Agora, substituindo (C.13) em (C.12), temos que

$$\pi(\lambda|\mathbf{z}, \boldsymbol{\beta}, \lambda) = C^{-1} K \exp \left\{ -\frac{1}{2\nu} (\lambda - m)^2 \right\}. \quad (\text{C.14})$$

Note que C e K não dependem de λ e o fator restante da distribuição (C.14) é o núcleo de uma distribuição normal com média m e variância ν . Portanto, a função (C.14) só será uma fdp se $C^{-1}K = (2\pi\nu)^{-1/2}$, o que implica que $C = \pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{w})$ é a fdp de uma distribuição normal n -variada dada por

$$\pi(\mathbf{z}|\boldsymbol{\beta}, \mathbf{w}) = \phi_n \left(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w}, \mathbf{I}_n + \tau\mathbf{w}\mathbf{w}^\top \right), \quad (\text{C.15})$$

e a distribuição condicional completa de λ é dada por

$$\lambda|\mathbf{z}, \mathbf{w}, \boldsymbol{\beta} \sim \mathcal{N}(m, \nu),$$

em que m e ν são dados em (C.13). ■

Proposição C.2 *Sob o modelo probito-assimétrico dado em (3.6), temos que a distribuição $\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda)$ é dada pela expressão (3.11).*

Prova: Utilizando o teorema de Bayes, temos que a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \mathbf{w}, \lambda\}$ é dada por

$$\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda) = C^{-1}\pi(\mathbf{z}, \mathbf{y}|\mathbf{w}, \lambda), \quad \text{sendo } C = \pi(\mathbf{y}|\mathbf{w}, \lambda). \quad (\text{C.16})$$

Temos ainda que $\pi(\mathbf{z}, \mathbf{y}|\mathbf{w}, \lambda) = \pi(\mathbf{z}|\mathbf{w}, \lambda)\pi(\mathbf{y}|\mathbf{z})$, onde $\pi(\mathbf{z}|\mathbf{w}, \lambda)$ e $\pi(\mathbf{y}|\mathbf{z})$ são dadas em (C.5) e (C.7), respectivamente. Logo, a distribuição (C.26) pode ser escrita como

$$\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda) = C^{-1}\phi_n\left(\mathbf{z}; \mathbf{X}\mathbf{b} - \lambda\mathbf{w}, \mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top\right)\text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.17})$$

Note que (C.17) só será uma fdp se $C = \bar{\Phi}_n(R(\mathbf{y}); \mathbf{X}\mathbf{b} - \lambda\mathbf{w}, \mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top)$, onde $\bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a função de distribuição acumulada na região $R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$ de uma distribuição normal n -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Deste modo, a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \mathbf{w}, \lambda\}$ é dada por

$$\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda \propto \mathcal{N}_n(\mathbf{X}\mathbf{b} - \lambda\mathbf{w}, \mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top)\text{Ind}(\mathbf{y}, \mathbf{z}), \quad (\text{C.18})$$

sendo $\text{Ind}(\mathbf{y}, \mathbf{z})$ a função indicadora dada em (C.7). Observe também que para $\lambda = 0$, a distribuição (C.18) se reduz à distribuição marginal de $\mathbf{z}$ *a posteriori* dada em (2.9) para modelo probito. ■

Proposição C.3 *Sob o modelo probito-assimétrico dado em (3.6), temos que a distribuição $\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \lambda)$ é dada pela expressão (3.13).*

Prova: Utilizando o teorema de Bayes, temos que a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \boldsymbol{\beta}, \lambda\}$ é dada por

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \lambda) = C^{-1}\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \lambda), \quad \text{sendo } C = \pi(\mathbf{y}|\boldsymbol{\beta}, \lambda). \quad (\text{C.19})$$

Considerando o modelo (3.6), temos que $z_i|w_i, \beta, \lambda \sim \mathcal{N}(\mathbf{x}_i^\top \beta - \lambda w_i, 1)$ e $w_i \sim \mathcal{HN}(0, 1)$. Assim, utilizando as propriedades da distribuição normal-assimétrica apresentadas no Apêndice D, segue que $z_i|\beta, \lambda \sim \mathcal{SN}(\mathbf{x}_i^\top \beta, 1 + \lambda^2, -\lambda)$. Deste modo, a distribuição do vetor aleatório $\mathbf{z}$ condicionado à $\{\beta, \lambda\}$ é dada por

$$\pi(\mathbf{z}|\beta, \lambda) = \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}_i^\top \beta}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \beta)\right]. \quad (\text{C.20})$$

Por outro lado, temos que $\pi(\mathbf{z}, \mathbf{y}|\beta, \lambda) = \pi(\mathbf{z}|\beta, \lambda)\pi(\mathbf{y}|\mathbf{z})$, onde $\pi(\mathbf{z}|\beta, \lambda)$ e $\pi(\mathbf{y}|\mathbf{z})$ são dadas em (C.20) e (C.7), respectivamente. Logo,

$$\pi(\mathbf{z}, \mathbf{y}|\beta, \lambda) = \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}_i^\top \beta}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \beta)\right] \prod_{i=1}^n \text{Ind}(y_i, z_i). \quad (\text{C.21})$$

Agora, substituindo (C.21) em (C.19), temos que

$$\begin{aligned} \pi(\mathbf{z}|\mathbf{y}, \beta, \lambda) &= C^{-1} \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}_i^\top \beta}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \beta)\right] \prod_{i=1}^n \text{Ind}(y_i, z_i) \\ &= C^{-1} \prod_{i=1}^n \left\{ \phi\left(\frac{z_i - \mathbf{x}_i^\top \beta}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \beta)\right] \text{Ind}(y_i, z_i) \right\}. \end{aligned}$$

Note que $C = \pi(\mathbf{y}|\beta, \lambda)$ não depende de $\mathbf{z}$ e o fator restante é o produto de distribuições normais-assimétricas univariadas. Portanto, os componentes z_i , $i = 1 \dots, n$, são variáveis aleatórias independentes com distribuição normal-assimétrica univariada truncada dada por

$$z_i|y_i, \beta, \lambda \propto \begin{cases} \mathcal{SN}(\mathbf{x}_i^\top \beta, 1 + \lambda^2, -\lambda)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{SN}(\mathbf{x}_i^\top \beta, 1 + \lambda^2, -\lambda)I(z_i \leq 0) & \text{se } y_i = 0. \end{cases}$$

■

Proposição C.4 *Sob o modelo probito-assimétrico dado em (3.6), temos que a distribuição $\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \beta)$ é dada pela expressão (3.14).*

Prova: Utilizando o teorema de Bayes, temos que a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \mathbf{w}, \beta\}$ é dada por

$$\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \beta) = C^{-1} \pi(\mathbf{z}, \mathbf{y}|\mathbf{w}, \beta), \quad \text{sendo } C = \pi(\mathbf{y}|\mathbf{w}, \beta). \quad (\text{C.22})$$

Temos também que $\pi(\mathbf{z}, \mathbf{y} | \mathbf{w}, \boldsymbol{\beta}) = \pi(\mathbf{z} | \mathbf{w}, \boldsymbol{\beta}) \pi(\mathbf{y} | \mathbf{z})$, onde $\pi(\mathbf{z} | \mathbf{w}, \boldsymbol{\beta})$ e $\pi(\mathbf{y} | \mathbf{z})$ são dadas em (C.15) e (C.7), respectivamente. Logo

$$\pi(\mathbf{z}, \mathbf{y} | \mathbf{w}, \boldsymbol{\beta}) = \phi_n \left(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w}, \mathbf{I}_n + \mathbf{w}\nu\mathbf{w}^\top \right) \text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.23})$$

Agora, substituindo (C.23) em (C.22), temos que

$$\pi(\mathbf{z} | \mathbf{y}, \mathbf{w}, \boldsymbol{\beta}) = C^{-1} \phi_n \left(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w}, \mathbf{I}_n + \mathbf{w}\nu\mathbf{w}^\top \right) \text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.24})$$

Note que (C.24) só será uma fdp se $C = \bar{\Phi}_n(R(\mathbf{y}); \mathbf{X}\mathbf{b}, \mathbf{I}_n + \mathbf{w}\nu\mathbf{w}^\top)$, onde $\bar{\Phi}_n$ e $R(\mathbf{y}, \mathbf{z})$ são dadas em (C.17). Portanto, a distribuição de $\mathbf{z}$ condicionada à $\{\mathbf{y}, \mathbf{w}, \boldsymbol{\beta}\}$ é dada por

$$\mathbf{z} | \mathbf{y}, \mathbf{w}, \boldsymbol{\beta} \propto \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w}, \mathbf{I}_n + \mathbf{w}\nu\mathbf{w}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (\text{C.25})$$

sendo $\text{Ind}(\mathbf{y}, \mathbf{z})$ a função indicadora dada em (C.7). ■

Proposição C.5 *Sob o modelo probito-assimétrico dado em (3.6), temos que a distribuição $\pi(\mathbf{z} | \mathbf{y}, \mathbf{w})$ é dada pela expressão (3.18).*

Observe que, dado $\mathbf{w}$, o modelo (3.6) se reduz ao modelo probito. Logo, a prova da Proposição C.5 é obtida de modo similar à prova da Proposição C.2, assumindo que $\lambda = 0$.

Proposição C.6 *Sob o modelo probito-assimétrico dado em (3.6), temos que as distribuições condicionais completas de $(\mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda)$ são dadas por (3.12).*

Prova: Utilizando o teorema de Bayes, temos que a distribuição de cada componente z_i condicionado à $\{\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\beta}, \lambda\}$ é dada por

$$\pi(z_i | \mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda) = C^{-1} \pi(\mathbf{z} | \mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda), \quad (\text{C.26})$$

onde $C = \pi(\mathbf{z}_{-i} | \mathbf{y}, \mathbf{w}, \lambda)$ e $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. Da Proposição C.3, temos que

$$\pi(\mathbf{z} | \mathbf{y}, \mathbf{w}, \lambda) \propto \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X}\mathbf{b} - \lambda\mathbf{w})]^\top (\mathbf{I}_n + \mathbf{X}\nu\mathbf{X}^\top)^{-1} [\mathbf{z} - (\mathbf{X}\mathbf{b} - \lambda\mathbf{w})] \right\},$$

com $\mathbf{z} \in R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$. Agora, usando o Lema C.3, temos que

$$\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda) \propto K \exp \left\{ -\frac{1-h_{ii}}{2} \left[z_i - \left(\frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 \right\} \text{Ind}(\mathbf{y}, \mathbf{z}),$$

sendo

$$\begin{aligned} K &= \exp \left\{ \frac{1}{2} \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (z_k - \mu_k) \left(h_{kj} - \frac{h_{ki}h_{ij}}{1-h_{ii}} \right) (z_j - \mu_j) \right\} \\ h_{ij} &= \mathbf{x}_i^\top (\boldsymbol{\nu}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_j \\ \mu_i &= \mathbf{x}_i^\top \mathbf{b} - \lambda w_i. \end{aligned}$$

Observe que K não depende de z_i . Portanto, podemos escrever a distribuição condicional completa de z_i como

$$\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda) \propto \exp \left\{ -\frac{1-h_{ii}}{2} \left[z_i - \left(\frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 \right\} \text{Ind}(y_i, z_i),$$

sendo $\text{Ind}(y_i, z_i) = I(z_i > 0)I(y_i = 1) + I(z_i \leq 0)I(y_i = 0)$. Note agora que a fdp $\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda)$ dada acima é proporcional ao núcleo de uma distribuição normal truncada. Assim, a distribuição condicional completa de cada componente z_i , $i = 1, \dots, n$ é dada por

$$z_i|\mathbf{z}_{-i}, y_i, \mathbf{w}, \lambda \propto \begin{cases} \mathcal{N}(m_i, \nu_i)I(z_i > 0) & \text{se } y_i = 1, \\ \mathcal{N}(m_i, \nu_i)I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (\text{C.27})$$

sendo $m_i = \mathbf{x}_i^\top \mathbf{b} - \lambda w_i + \frac{1}{1-h_{ii}} \sum_{\substack{k=1 \\ k \neq i}}^n h_{ik}(z_k - \mathbf{x}_k^\top \mathbf{b} + \lambda w_k)$ e $\nu_i = \frac{1}{1-h_{ii}}$.

Por outro lado, o parâmetro de localização m_i pode ser reescrito em função da média *a posteriori* de $\boldsymbol{\beta}$, ou seja, em função de $\mathbf{B} = \mathbf{V} [\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z} + \lambda \mathbf{w})]$, sendo $\mathbf{V} = (\boldsymbol{\nu}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1}$. Essa representação fornece uma estrutura conveniente para o uso do algoritmo de Gibbs. Assim, escrevendo $h_{ij} = \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j$, temos que

$$\begin{aligned} m_i &= \frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j (z_j - \mathbf{x}_j^\top \mathbf{b} + \lambda w_j) + \mathbf{x}_i^\top \mathbf{b} - \lambda w_i \\ &= \frac{1}{1-h_{ii}} \left[\sum_{j=1}^n \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j (z_j - \mathbf{x}_j^\top \mathbf{b} + \lambda w_j) - \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_i (z_i - \mathbf{x}_i^\top \mathbf{b} + \lambda w_i) \right] + \mathbf{x}_i^\top \mathbf{b} - \lambda w_i. \end{aligned}$$

Agora, usando o fato que $\sum_{j=1}^n \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j (z_j - \mathbf{x}_j^\top \mathbf{b} + \lambda w_j) = \mathbf{x}_i^\top \mathbf{V} \mathbf{X} (\mathbf{z} - \mathbf{X} \mathbf{b} + \lambda \mathbf{w}) = \mathbf{x}_i^\top \mathbf{B} - \mathbf{x}_i^\top \mathbf{b}$, temos que

$$\begin{aligned} m_i &= \frac{1}{1 - h_i} \left[\mathbf{x}_i^\top \mathbf{B} - \mathbf{x}_i^\top \mathbf{b} - h_i (z_i - \mathbf{x}_i^\top \mathbf{b} + \lambda w_i) \right] + \mathbf{x}_i^\top \mathbf{b} - \lambda w_i \\ m_i &= \mathbf{x}_i^\top \mathbf{B} - \lambda w_i - \frac{h_i}{1 - h_i} \left[z_i - (\mathbf{x}_i^\top \mathbf{B} - \lambda w_i) \right]. \end{aligned} \quad (\text{C.28})$$

Para o uso do algoritmo de Gibbs, é interessante que não seja necessário calcular o vetor $\mathbf{B}$ diretamente a cada atualização de algum z_i . Deste modo, fazemos a seguinte atualização de $\mathbf{B}$,

$$\begin{aligned} \mathbf{B} &= \mathbf{V} \left[\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z} + \lambda \mathbf{w}) \right] = \mathbf{V} \left\{ \boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top [\mathbf{z}^{ant} + \mathbf{1}_i (z_i - z_i^{ant}) + \lambda \mathbf{w}] \right\} \\ \mathbf{B} &= \mathbf{V} \left[\boldsymbol{\nu}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z}^{ant} + \lambda \mathbf{w}) \right] + \mathbf{V} \mathbf{X}^\top \mathbf{1}_i (z_i - z_i^{ant}) \\ \mathbf{B} &= \mathbf{B}^{ant} + \mathbf{s}_i (z_i - z_i^{ant}), \end{aligned} \quad (\text{C.29})$$

onde $\mathbf{B}^{ant}$ e z_i^{ant} denotam, respectivamente, os valores armazenados das atualizações anteriores de $\mathbf{B}$ e z_i ; $\mathbf{s}_i$ denota o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V} \mathbf{X}^\top$, e $\mathbf{1}_i$ denota um vetor de dimensão $(n \times 1)$ com o i -ésimo elemento igual a um e todos os elementos restantes iguais a zero.

Observe também que para $\lambda = 0$, as distribuições condicionais completas dadas em (C.27) se reduzem às distribuições condicionais completas de $(\mathbf{z}|\mathbf{y})$ para o modelo probito dadas em (2.10). ■

Proposição C.7 *Sob o modelo probito-assimétrico dado em (3.6), temos que as distribuições condicionais completas de $(\mathbf{z}|\mathbf{y}, \mathbf{w}, \boldsymbol{\beta})$ são dadas por (3.15).*

Prova: Utilizando o teorema de Bayes, temos que a distribuição de cada componente z_i condicionado à $\{\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\beta}, \lambda\}$ é dada por

$$\pi(z_i | \mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \boldsymbol{\beta}) = C^{-1} \pi(\mathbf{z} | \mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \boldsymbol{\beta}), \quad (\text{C.30})$$

onde $C = \pi(\mathbf{z}_{-i} | \mathbf{y}, \mathbf{w}, \boldsymbol{\beta})$ e $\mathbf{z}_{-i}$ denota o vetor de variáveis $\mathbf{z}$ com a i -ésima variável removida. Da Proposição C.4, temos que

$$\pi(\mathbf{z} | \mathbf{y}, \mathbf{w}, \boldsymbol{\beta}) \propto \exp \left\{ -\frac{1}{2} [\mathbf{z} - (\mathbf{X} \boldsymbol{\beta} - \alpha \mathbf{w})]^\top (\mathbf{I}_n + \tau \mathbf{w} \mathbf{w}^\top)^{-1} [\mathbf{z} - (\mathbf{X} \boldsymbol{\beta} - \alpha \mathbf{w})] \right\},$$

com $\mathbf{z} \in R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, z_2, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$. Usando o Lema C.3, temos que

$$\pi(\mathbf{z}|\mathbf{y}, \mathbf{w}, \lambda) \propto K \exp \left\{ -\frac{1-h_{ii}}{2} \left[z_i - \left(\frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 \right\} \text{Ind}(\mathbf{y}, \mathbf{z}),$$

sendo

$$\begin{aligned} K &= \exp \left\{ \frac{1}{2} \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (z_k - \mu_k) \left(h_{kj} - \frac{h_{ki}h_{ij}}{1-h_{ii}} \right) (z_j - \mu_j) \right\} \\ h_{ij} &= \mathbf{w}_i^\top \left(\tau^{-1} + \mathbf{w}^\top \mathbf{w} \right)^{-1} \mathbf{w}_j \\ \mu_i &= \mathbf{x}_i^\top \boldsymbol{\beta} - \alpha w_i. \end{aligned}$$

Note que K não depende de z_i . Assim, podemos escrever a distribuição condicional completa de z_i como

$$\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda) \propto \exp \left\{ -\frac{1-h_{ii}}{2} \left[z_i - \left(\frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}(z_j - \mu_j) + \mu_i \right) \right]^2 \right\} \text{Ind}(y_i, z_i),$$

sendo $\text{Ind}(y_i, z_i) = I(z_i > 0)I(y_i = 1) + I(z_i \leq 0)I(y_i = 0)$. Observe agora que a fdp $\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \mathbf{w}, \lambda)$ dada acima é proporcional ao núcleo de uma distribuição normal truncada. Portanto, a distribuição condicional completa de cada componente z_i é dada por

$$z_i|\mathbf{z}_{-i}, y_i, \mathbf{w}, \lambda \propto \begin{cases} N(m_i, \nu_i)I(z_i > 0) & \text{se } y_i = 0, \\ N(m_i, \nu_i)I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (\text{C.31})$$

sendo $m_i = \mathbf{x}_i^\top \boldsymbol{\beta} - \alpha w_i + \frac{1}{1-h_{ii}} \sum_{\substack{k=1 \\ k \neq i}}^n h_{ik}(z_k - \mathbf{x}_k^\top \boldsymbol{\beta} + \alpha w_k)$ e $\nu_i = \frac{1}{1-h_{ii}}$.

Por outro lado, o parâmetro de localização m_i pode ser reescrito em função da média *a posteriori* de λ , ou seja, em função de $m = \nu [\tau^{-1}\alpha - \mathbf{w}^\top(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})]$, sendo $\nu = (\mathbf{w}^\top \mathbf{w} + \tau^2)^{-1}$. Essa representação fornece uma estrutura conveniente para o uso do algoritmo de Gibbs. Portanto, escrevendo $h_{ij} = w_i \nu w_j$, temos que

$$\begin{aligned} m_i &= \frac{1}{1-h_{ii}} \sum_{\substack{j=1 \\ j \neq i}}^n w_i \nu w_j (z_j - \mathbf{x}_j^\top \boldsymbol{\beta} + \alpha w_j) + \mathbf{x}_i^\top \boldsymbol{\beta} - \alpha w_i \\ &= \frac{1}{1-h_{ii}} \left[\sum_{j=1}^n w_i \nu w_j (z_j - \mathbf{x}_j^\top \boldsymbol{\beta} + \alpha w_j) - w_i \nu w_i (z_i - \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha w_i) \right] + \mathbf{x}_i^\top \mathbf{b} - \lambda w_i \end{aligned}$$

Agora, usando o fato que $\sum_{j=1}^n w_i \nu \mathbf{x}_j (z_j - \mathbf{x}_j^\top \boldsymbol{\beta} + \alpha w_j) = w_i \nu \mathbf{w} (z - \mathbf{X} \boldsymbol{\beta} + \alpha \mathbf{w}) = w_i m - w_i \alpha$, temos

$$\begin{aligned} m_i &= \frac{1}{1 - h_i} \left[w_i m - w_i \alpha - h_i (z_i - \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha w_i) \right] + \mathbf{x}_i^\top \boldsymbol{\beta} - \alpha w_i \\ m_i &= w_i m - \alpha w_i - \frac{h_i}{1 - h_i} [z_i - (w_i m - \alpha w_i)]. \end{aligned}$$

Para o uso do algoritmo de Gibbs, é interessante que não seja necessário calcular m diretamente a cada atualização de algum z_i . Deste modo, fazemos a seguinte atualização de m ,

$$\begin{aligned} m &= \nu \left[\tau^{-1} \alpha - \mathbf{w}^\top (z - \mathbf{X} \boldsymbol{\beta}) \right] = \nu \left\{ \tau^{-1} \alpha - \mathbf{w}^\top [\mathbf{z}^{ant} + \mathbf{1}_i (z_i - z_i^{ant}) - \mathbf{X} \boldsymbol{\beta}] \right\} \\ m &= \nu \left[\tau^{-1} \alpha - \mathbf{w}^\top (\mathbf{z}^{ant} - \mathbf{X} \boldsymbol{\beta}) \right] + \nu \mathbf{w}^\top \mathbf{1}_i (z_i - z_i^{ant}) \\ m &= m^{ant} + s_i (z_i - z_i^{ant}), \end{aligned}$$

sendo m^{ant} e z_i^{ant} , respectivamente, os valores armazenados das atualizações anteriores de m e z_i ; $s_i = \nu w_i$ e $\mathbf{1}_i$ denota um vetor de dimensão $(n \times 1)$ com o i -ésimo elemento igual a um e todos os elementos restantes iguais a zero. ■

Proposição C.8 *Sob o modelo probito-assimétrico dado em (3.6), temos que as distribuições condicionais completas de $(\mathbf{z}|\mathbf{y}, \mathbf{w})$ são dadas por (3.19).*

Observe que, dado $\mathbf{w}$, o modelo (3.6) se reduz ao modelo probito. Logo, a prova da Proposição C.8 é obtida de modo similar à prova da Proposição C.6, assumindo que $\lambda = 0$.

Apêndice D

Algumas distribuições utilizadas

D.1 Distribuição normal-assimétrica

A distribuição normal-assimétrica foi introduzida formalmente por Azzalini (1985), e suas propriedades são apresentadas em diversos trabalhos, entre eles, Azzalini (1985) e Branco e Arellano-Valle (2004).

Uma variável aleatória Z tem distribuição normal-assimétrica (*skew-normal*), com vetor de parâmetros $\theta = (\mu, \sigma^2, \lambda)$, em que $\mu \in \mathbb{R}$ é o parâmetro de localização, $\sigma^2 > 0$ é o parâmetro de escala, e $\lambda \in \mathbb{R}$ é o parâmetro de forma, se sua função densidade é dada por

$$f(z) = \frac{2}{\sigma} \phi\left(\frac{z - \mu}{\sigma}\right) \Phi\left(\lambda \frac{z - \mu}{\sigma}\right), \quad \text{com } z \in \mathbb{R}. \quad (\text{D.1})$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ denotam, respectivamente, as funções densidade e de distribuição de uma variável normal padrão. Esta distribuição será denotada por $\mathcal{SN}(\mu, \sigma^2, \lambda)$ e sua função de distribuição por $\Phi_{\mathcal{SN}}(\cdot; \mu, \sigma^2, \lambda)$.

Um caso especial interessante da distribuição normal-assimétrica ocorre quando o vetor de parâmetros $\theta = (0, 1 + \lambda^2, \lambda)$, neste caso, a densidade de Z em (D.1) se reduz a densidade da normal-assimétrica dada por Chen *et al* (1999), que é um caso particular da normal-assimétrica multivariada proposta em Sahu, Dey e Branco (2003). Chamaremos essa distribuição de normal-assimétrica *CDS*, e denotaremos sua função de distribuição acumulada (fda) por $\Phi_{CDS}(\cdot; \lambda)$, que tem função densidade dada por

$$f(z) = \frac{2}{\sqrt{1 + \lambda^2}} \phi\left(\frac{z}{\sqrt{1 + \lambda^2}}\right) \Phi\left(\lambda \frac{z}{\sqrt{1 + \lambda^2}}\right), \quad \text{com } z \in \mathbb{R}. \quad (\text{D.2})$$

A distribuição normal-assimétrica dada em (D.1) tem algumas propriedades interessantes. Dentre essas, as que foram úteis para o nosso trabalho são apresentadas a seguir.

Propriedades

P1. Quando $\lambda = 0$, obtemos a densidade da distribuição $\mathcal{N}(\mu, \sigma^2)$.

P2. Quando $\mu = 0$, $\sigma^2 = 1$ e $\lambda \rightarrow \infty$, obtemos a densidade da distribuição *Half-Normal* (ou normal positiva), a qual é dada por

$$g(w) = \frac{2}{\sqrt{2\pi}} e^{-w^2/2}, \quad \text{com } w > 0.$$

Essa distribuição será denotada por $\mathcal{HN}(1, 0)$.

P3. Sejam $\epsilon \sim \mathcal{N}(0, 1)$ e $w \sim \mathcal{HN}(0, 1)$ variáveis aleatórias independentes, e $\lambda \in \mathbb{R}$, então

$$Z = \mu + \lambda w + \epsilon \sim \mathcal{SN}(\mu, 1 + \lambda^2, \lambda).$$

P4. Se $Z|W = w \sim \mathcal{N}(\mu + \lambda w, 1)$, $w \sim \mathcal{HN}(0, 1)$ e $\lambda \in \mathbb{R}$, então

$$Z \sim \mathcal{SN}(\mu, 1 + \lambda^2, \lambda).$$

P5. Se $Z \sim \mathcal{SN}(\mu, \sigma^2, \lambda)$ e $a, b \in \mathbb{R}$, então

$$Z_2 = a + bZ \sim \mathcal{SN}(a + b\mu, b^2\sigma^2, \text{sinal}(b), \lambda).$$

P6. Se $Z \sim \mathcal{SN}(\mu, \sigma^2, \lambda)$, então $\frac{(Z - \mu)^2}{\sigma^2} \sim \chi_1^2$, onde χ_n^2 denota uma distribuição qui-quadrado com n graus de liberdade.

P7. $\Phi_{\mathcal{SN}}(z; \mu, \sigma^2, \lambda) = 1 - \Phi_{\mathcal{SN}}(-z; -\mu, \sigma^2, -\lambda)$.

P8. A fda de uma distribuição normal-assimétrica pode ser obtida através de uma distribuição normal bivariada da seguinte forma

$$\Phi_{\mathcal{SN}}(z; \mu, \sigma^2, \lambda) = 2\Phi_2 \left(\begin{pmatrix} \frac{z-\mu}{\sigma} \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -d \\ d & 1 \end{pmatrix} \right), \quad \text{sendo } d = \frac{\lambda}{\sqrt{1 + \lambda^2}},$$

onde $\Phi_2(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a fda da distribuição normal bivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$.

A relação entre as fda's das distribuições normal-assimétrica e normal bivariada é útil para implementação computacional.

D.2 Distribuição gaussiana-inversa

A distribuição gaussiana-inversa, denotada por GI , tem atraído uma considerável atenção nas últimas décadas, principalmente depois do trabalho de Jorgensen (1982). A distribuição gaussiana-inversa tem sido utilizada para explicar o movimento browniano, e mais recentemente, na teoria de modelos lineares generalizados (McCullagh e Nelder, 1983). Para detalhes, ver Chhikara e Falks (1989) e Jorgensen (1982).

Uma variável aleatória X tem distribuição gaussiana-inversa com parâmetros μ e λ , se sua função densidade de probabilidade é dada por

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad \text{com } x > 0,$$

sendo $\mu > 0$ e $\lambda > 0$ os parâmetros de localização e de forma, respectivamente. Essa distribuição é unimodal com todos os momentos finitos, ver Chhikara e Falks (1989).

Em Devroye (1986) e Chhikara e Falks (1989) é apresentado o método de *Michael, Schucany e Haas* para amostrar de uma distribuição gaussiana-inversa usando transformação com múltiplas raízes (Devroye, 1986, p. 145-149).

Se $X \sim IG(\mu, \lambda)$, então $Y = \frac{\lambda(X - \mu)^2}{2\mu^2 X} \sim \chi_1^2$. A transformação inversa $X = g^{-1}(Y)$ tem duas raízes, X_1 e X_2 , sendo

$$X_1 = \mu + \frac{\mu^2 Y}{2\lambda} - \frac{\mu}{2\lambda} \sqrt{4\mu\lambda Y + \mu^2 Y^2} \quad \text{e} \quad X_2 = \frac{\mu^2}{X_1}. \quad (\text{D.3})$$

Dado o valor Y gerado de χ_1^2 . A raiz X_1 é escolhida com probabilidade $\mu/(\mu + X_1)$ e a outra raiz com probabilidade $X_1/(\mu + X_1)$. O procedimento para gerar uma variável GI é dado a seguir:

- i) Gera-se uma variável aleatória Y com distribuição qui-quadrado com 1 grau de liberdade;
- ii) Para cada valor gerado no passo (i), calcula-se X_1 utilizando a expressão (D.3);
- iii) Gera-se uma variável aleatória U com distribuição uniforme no intervalo $(0, 1)$;
- iv) Se $U \leq \mu/(\mu + X_1)$, aceita-se X_1 como uma observação da variável aleatória GI , caso contrário $X_2 = \mu^2/X_1$ é aceita.

D.3 Distribuição gaussiana-inversa generalizada

A distribuição gaussiana-inversa generalizada, denotada por GIG , foi amplamente estudada por Jorgensen (1982). A família GIG tem tido sucesso recentemente devido suas aplicações em modelagem. Silva *et al* (2006) aplicaram suas extensões em modelos de volatilidade.

Uma variável aleatória X tem distribuição gaussiana-inversa generalizada com parâmetros λ , ψ e χ , denotada por $GIG(\lambda, \psi, \chi)$, se sua função densidade de probabilidade é dada por

$$f(x) = \frac{\left(\frac{\psi}{\chi}\right)^2}{2K_\lambda(\sqrt{\psi\chi})} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right\}, \quad \text{com } x > 0,$$

sendo $\lambda \in \mathbb{R}$, $\chi > 0$ e $\psi > 0$, os parâmetros da distribuição, e K_λ a função modificada de Bessel de terceira espécie indexada por λ .

A função $K_\lambda(w)$ pode ser encontrada na literatura com duas representações,

$$K_\lambda(w) = \frac{1}{2} \int_{-\infty}^{\infty} \cosh(\lambda u) e^{-w \cosh(u)} du \quad \text{ou} \quad K_\nu(w) = \frac{1}{2} \int_0^{\infty} u^{\nu-1} \exp\left\{-\frac{1}{2}w(u + u^{-1})\right\} du,$$

Essas são as representações utilizadas em Devroye (1986) e Jorgensen (1982), respectivamente.

A função de Bessel K_ν satisfaz as relações

$$K_\nu(w) = K_{-\nu}(w),$$

e para $\nu = n + 1/2$ e $n = 0, 1, 2, \dots$, temos

$$K_{n+\frac{1}{2}}(w) = \sqrt{\frac{\pi}{2w}} e^{-w} \left(1 + \sum_{i=1}^n \frac{(n+1)!}{(n-1)!i!} (2w)^{-i}\right).$$

Usando as relações acima, temos que $K_{-\frac{1}{2}}(w) = K_{\frac{1}{2}}(w) = \sqrt{\frac{\pi}{2w}} e^{-w}$. Outras propriedades da função de Bessel podem ser vistas em Silva *et al* (2006).

Os parâmetros da distribuição gaussiana-inversa generalizada têm as seguintes relações de dependência: (i) $\chi \geq 0$, $\psi > 0$ se $\lambda > 0$; (ii) $\chi > 0$, $\psi > 0$ se $\lambda = 0$; e (iii) $\chi > 0$, $\psi \geq 0$

se $\lambda < 0$. Então, condicionada à dependência existente entre os parâmetros, podemos apresentar alguns casos especiais desta família de distribuições e algumas relações entre distribuições.

Alguns casos especiais da distribuição *GIG*:

- C1. Quando $\lambda > 0$ e $\chi \rightarrow 0$, obtemos a densidade de uma gama.
- C2. Quando $\lambda < 0$ e $\psi \rightarrow 0$, obtemos a densidade de uma gama-inversa.
- C3. Quando $\lambda = -\frac{1}{2}$, obtemos a densidade de uma gaussiana-inversa.

Algumas relações entre distribuições

- R1. $GIG(\lambda, \psi, \chi) = \frac{1}{c} GIG(\lambda, \frac{\psi}{c}, c\chi)$ para todo $c > 0$. Em particular,

$$GIG(\lambda, \psi, \chi) = \sqrt{\frac{\chi}{\psi}} GIG(\lambda, \sqrt{\psi\chi}, \sqrt{\psi\chi}).$$

- R2. $GIG(\lambda, \psi, \chi) = \frac{1}{GIG(-\lambda, \psi, \chi)}.$

D.4 Distribuição Kolmogorov-Smirnov

Uma variável aleatória X tem distribuição Kolmogorov-Smirnov assintótica (*KS*) se sua função de distribuição é dada por

$$F(x) = \sum_{n=-\infty}^{\infty} (-1)^n \exp(-2n^2 x^2), \quad \text{com } x > 0,$$

e sua função densidade de probabilidade por

$$f(x) = 8 \sum_{n=1}^{\infty} (-1)^{n+1} n^2 x \exp(-2n^2 x^2), \quad \text{com } x > 0.$$

Outra forma de escrever a fdp da *KS* é dada por

$$f(x) = \frac{\sqrt{2\pi}}{x} \sum_{n=1}^{\infty} \left[\frac{(2n-1)^2 \pi^2}{4x^3} - \frac{1}{x} \right] \exp \left\{ \frac{(2n-1)^2 \pi^2}{8x^2} \right\}, \quad \text{com } x > 0,$$

e a função de distribuição

$$F(x) = \sum_{n=1}^{\infty} \exp \left\{ \frac{(2n-1)^2 \pi^2}{8x^2} \right\}, \quad \text{com } x > 0.$$

Referências Bibliográficas

- [1] Albert, J. H. e Chib, S. (1993). Bayesian Analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679.
- [2] Albert, J. H. e Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika*, **82**, 747-759.
- [3] Andrews, D. F. e Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of Royal Statistical Society, Series B*, **36**, 99-102.
- [4] Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, **68**, 357-363.
- [5] Azzalini A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal Statistical*, **12**, 171-178.
- [6] Azzalini, A. e Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of Royal Statistical Society, Series B*, **65**, 367-389.
- [7] Azzalini, A. (2006). R package “sn”: The skew-normal and skew-t distributions (version 0.4-0). URL <http://azzalini.stat.unipd.it/SN>.
- [8] Atkinson, A. C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal of Scientific and Statistical Computations*, **3**, 502-515.
- [9] Basu, S. e Mukhopadhyay, S. (2000). Binary Response Regression with Normal Scale Mixtures Links, em *Generalized Linear Models: A Bayesian Perspective*, eds. D.K. Dey, S. K. Ghosh, e B. K. Mallick, New York: Marcel Dekker.

-
- [10] Bazán, J. L. (2005). *Uma Família de Modelos de Resposta ao Item Normal Assimétrica*. Tese de Doutorado, IME-USP, São Paulo.
- [11] Bazán, J. L., Bolfanini, H. e Branco, M. D. (2006). A generalized skew probit class link for binary regression. RT-MAE 2006-05 .
- [12] Bazán, J. L., Branco, M. D. e Bolfanini, H. (2005). A skew item reponse model. RT-MAE 2005-06. *Bayesian Analysis*, **1**(4), 861-892 .
- [13] Bliss, C. I. (1935). The calculation of the dose-mortality curve. *Annals Applied Biology*, **22**, 134-167.
- [14] Branco, M. D. e Arellano-Valle, R. (2004). Distribuições Elípticas Assimétricas.16o. SINAPE - ABE. 54p
- [15] 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.
- [16] Chen, M-H. (2004). The Skewed Link Models for Categorical Response Data, em *Skew-elliptical distributions and their applications: a journey beyond normality*, ed. Marc. G. Genton, Boca Raton: Chapman and Hall/CRC.
- [17] 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.
- [18] Chhikara, R. S. e Folks, J. L. (1989). *The Inverse Gaussian Distribution*. New York: Marcel Dekker. 213p
- [19] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer. 843p.
- [20] Holmes, C. C. e Held, L. (2006). *Bayesian auxiliary variable models for binary and multinomial regression*. *Bayesian Analysis*, **1**(1), 145-168.
- [21] 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.

- [22] Jorgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution, *Lecture Notes in Statistics*, No. 9, New York: Springer-Verlag.
- [23] McCullagh, P. e Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall, 2a edição. 511p.
- [24] Neal, R. M. (1993) *Probabilistic Inference Using Markov Chain Monte Carlo Methods*, Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto, 144p. URL <http://www.cs.toronto.edu/~radford/papers-online.html>.
- [25] Paulino, C., Turkman, M. e Murteira, B. (2003) *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa. 446p.
- [26] Plummer, M., Best, N., Cowles, K. e Vines, K. (2006). coda: Output analysis and diagnostics for MCMC. *R package version 0.10-7*.
- [27] Prentice, R. L. (1976). A Generalization of the Probit and Logit Methods for Dose-Response Curves. *Biometrika*, **32**, 761-768.
- [28] R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [29] Sahu, S. K., Dey, D. K. e Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **29**, 217-232.
- [30] Sen, P. K. e Singer, J. M. (1993) *Large Sample Methods in Statistics: An Introduction with Applications*, Chapman and Hall, London. 382p.
- [31] Silva, R. S., Lopes, H. F. e Migon, H. S. (2006). The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models. *Brazilian Journal of Probability and Statistics*, **20**, 67-91.
- [32] Stukel, T. (1988). Generalized Logistic Models. *Journal of the American Statistical Association*, **83**, 426-431.