
Regressão binária nas abordagens clássica e bayesiana

Amélia Milene Correia Fernandes

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Amélia Milene Correia Fernandes

Regressão binária nas abordagens clássica e bayesiana

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestra em Estatística – Interinstitucional de Pós-Graduação em Estatística. VERSÃO REVISADA

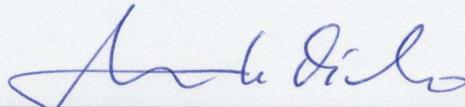
Área de Concentração: Estatística

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

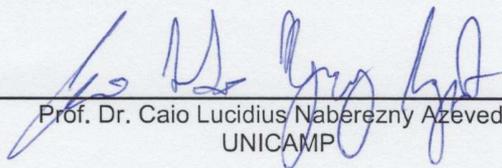
**USP – São Carlos
Fevereiro de 2017**

Folha de Aprovação

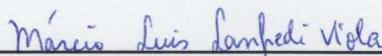
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Amélia Milene Correia Fernandes, realizada em 16/12/2016:



Prof. Dr. Marinho Gomes de Andrade Filho
USP



Prof. Dr. Caio Lucidius Naberezny Azevedo
UNICAMP



Prof. Dr. Marcio Luis Lanfredi Viola
UFSCar

Amélia Milene Correia Fernandes

Binary regression in the classical and Bayesian approaches

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação - ICMC-USP and to the Departamento de Estatística – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP. FINAL VERSION

Concentration Area: Statistics

Advisor: Prof. Dr. Marinho Gomes de Andrade Filho

**USP – São Carlos
February 2017**

*Dedico este trabalho à
minha família.*

Agradecimentos

Agradeço primeiramente a Deus por ter iluminado o meu caminho.

Aos meus pais, Angelina e Gabriel, pela confiança, apoio, incentivo, carinho e amor que recebi durante todo o curso do mestrado. Mesmo longe, vocês sempre estiveram presente me amparando nos momentos difíceis e fazendo o possível para que esse sonho se realizasse. Muito obrigada pela compreensão por estar tanto tempo longe.

Ao meu orientador, Prof. Marinho Andrade, pela atenção, paciência e oportunidade de trabalhar ao seu lado e por ser o maior incentivador na superação dos meus limites. Muito obrigada por ser referência de profissional. Me tornarei uma profissional melhor com os seus conselhos e exemplos.

Aos professores Caio Azevedo, Carlos Diniz e Márcio Viola membros da banca examinadora, pelas sugestões feitas.

Aos meus irmãos, Frederico, Maria Neusa, Neusa Dilce, Henriqueta e Henrique, por serem pessoas maravilhosas e por sempre me apoiarem nas minhas decisões.

Aos meus sobrinhos e as minhas sobrinhas por alegrarem a minha vida. Vocês são crianças maravilhosas, sou sortuda e grata por ter vocês presente na minha vida.

À Jéssica, pela amizade, compreensão e apoio que ela sempre me deu de forma tão gentil. Obrigada pela sua generosidade.

Aos meus amigos e colegas do laboratório, em especial, Mary, Thales, Juan, Helio, Ricardo, Dantes, Fabiano, Rafael e Brenno. Obrigada pela força, paciência e pela troca de experiência.

Ao departamento da Estatística de UFSCar e do ICMC e aos professores que contribuíram para o meu aprendizado.

Aos meus tios, tias, primos e primas, especialmente tia Maria Eduarda, prima Rose, tio Daniel pelo incentivo e apoio nessa jornada.

Por fim, a todos que me incentivaram, apoiaram e aconselharam. Agradeço por vocês fazarem parte da minha vida.

Resumo

Este trabalho tem como objetivo estudar o modelo de regressão binária nas abordagens clássica e bayesiana utilizando as funções de ligações probito, logito, complemento log-log, transformação box-cox e probito-assimétrico. Na abordagem clássica apresentamos as suposições e o procedimento para ajustar o modelo de regressão e verificamos a precisão dos parâmetros estimados, construindo intervalos de confiança e testes de hipóteses. Enquanto que, na inferência bayesiana fizemos um estudo comparativo utilizando duas metodologias. Na primeira metodologia consideramos densidades *a priori* não informativas e utilizamos o algoritmo Metropolis-Hastings para ajustar o modelo. Na segunda metodologia utilizamos variáveis auxiliares para obter a distribuição *a posteriori* conhecida, facilitando a implementação do algoritmo do Amostrador de Gibbs. No entanto, a introdução destas variáveis auxiliares podem gerar valores correlacionados, o que leva à necessidade de se utilizar o agrupamento das quantidades desconhecidas em blocos para reduzir a autocorrelação.

Através do estudo de simulação mostramos que na inferência clássica podemos usar os critérios AIC e BIC para escolher o melhor modelo e avaliamos se o percentual de cobertura do intervalo de confiança assintótica está de acordo com o esperado na teoria assintótica. Na inferência bayesiana constatamos que o uso de variáveis auxiliares resulta em um algoritmo mais eficiente segundo os critérios: erro quadrático médio (EQM), erro percentual absoluto médio (MAPE) e erro percentual absoluto médio simétrico (SMAPE).

Como ilustração apresentamos duas aplicações com dados reais. Na primeira, consideramos um conjunto de dados da variação do Ibovespa e a variação do valor diário do fechamento da cotação do dólar no período de 2013 a 2016. Na segunda aplicação, trabalhamos com um conjunto de dados educacionais (INEP-2013), focando nos estudos das variáveis que influenciam a aprovação do aluno.

Palavras-chave: *Modelo de regressão binária, inferência clássica, inferência bayesiana, variável auxiliar, função de ligação.*

Abstract

The objective of this work is to study the binary regression model under the frequentist and Bayesian approaches using the probit, logit, log-log complement, Box-Cox transformation and skew-probit as link functions. In the classical approach we presented assumptions and procedures used in the regression modeling. We verified the accuracy of the estimated parameters by building confidence intervals and conducting hypothesis tests. In the Bayesian approach we made a comparative study using two methodologies. For the first methodology, we considered non-informative prior distributions and the Metropolis-Hastings algorithm to estimate the model. In the second methodology we used auxiliary variables to obtain the known a posteriori distribution, allowing the use of the Gibbs Sampler algorithm. However, the introduction of these auxiliary variables can generate correlated values and needs the use of clustering of unknown quantities in blocks to reduce the autocorrelation. In the simulation study we used the AIC and BIC information criteria to select the most appropriate model and we evaluated whether the coverage probabilities of the confidence interval is in agreement with that expected by the asymptotic theory. In Bayesian approach we found that the inclusion of auxiliary variables in the model results in a more efficient algorithm according to the MSE, MAPE and SMAPE criteria. In this work we also present applications to two real datasets. The first dataset used is the variation of the Ibovespa and variation of the daily value of the American dollar at the time of closing the 2013 to 2016. The second dataset, used is an educational data set

(INEP-2013), where we are interested in studying the factors that influence the approval of the student.

Key words: *Binary regression model, classical inference, Bayesian inference, link function.*

Sumário

1	Introdução	1
1.1	Regressão Binária	2
1.2	Funções de Ligações	3
1.2.1	Logito	3
1.2.2	Probita	3
1.2.3	Complemento log-log	3
1.2.4	Transformação box-cox	4
1.2.5	Probita-assimétrico	5
1.3	Objetivos e Organização da Dissertação	6
2	Abordagem Clássica	9
2.1	Estimação dos Parâmetros	9
2.1.1	Estimação para as Funções de Ligações Assimétricas	13
2.2	Intervalo de Confiança	14
2.3	Testes de Hipóteses	15
2.3.1	Teste de razão de verossimilhanças	15
2.3.2	Teste Wald	15
2.3.3	Teste Escore	15
2.4	Função Desvio	16
2.5	Diagnósticos Clássicos de Regressão	17
2.5.1	Resíduos de Pearson	17
2.5.2	Resíduos de Anscombe	18
2.5.3	Resíduos de Pearson Studentizados	18
2.5.4	Componentes do Desvio	18
2.5.5	Componentes do Desvio Studentizados	19
2.5.6	Distância de Cook	19

2.6	Critério de Seleção de Modelos	19
2.6.1	Critério de Akaike - AIC	20
2.6.2	Critério de Informação Bayesiano - BIC	20
2.7	Propriedades dos Estimadores	21
2.8	Considerações Finais	21
3	Abordagem Bayesiana	23
3.1	Inferência Bayesiana Sem Variáveis Auxiliares	23
3.2	Inferência Bayesiana com Variáveis Auxiliares	25
3.2.1	Regressão Probita	25
3.2.2	Regressão Logística	27
3.2.3	Modelo Probita-Assimétrico	30
3.3	Algoritmos de simulação	32
3.3.1	Atualização conjunta de $(\mathbf{z}, \boldsymbol{\beta})$	32
3.3.2	Atualização conjunta de $(\mathbf{z}, \boldsymbol{\omega})$	34
3.3.3	Atualização conjunta de $(\mathbf{z}, \boldsymbol{\beta}, \lambda)$	34
3.4	Identificação de Pontos Influentes	36
3.5	Critérios Para Comparação de Modelos	38
3.6	Medidas de Eficiência	38
3.7	Critério de Convergência	40
3.8	Considerações Finais	41
4	Resultados de Simulações	43
4.1	Inferência Clássica	44
4.2	Inferência bayesiana	46
5	Aplicações	59
5.1	Aplicação aos Dados do Ibovespa	59
5.2	Aplicação aos Dados Educacionais	62
6	Considerações Finais	69
	Referências Bibliográficas	70
A	Pseudo-Códigos	71
A.1	Procedimento para amostragem de Probita	72
A.2	Procedimentos para amostragem no modelo logístico	73
A.3	Procedimentos para amostragem no modelo probita-assimétrico	78

B	Método para amostrador da variável misturadora no modelo logístico	81
C	Provas das distribuições da função Probit-assimétrico	85
D	Pseudo-Códigos	95
D.1	Distribuição normal-assimétrica	95
D.2	Distribuição guassiana-inversa	96
D.3	Distribuição guassiana-inversa generalizada	97
D.4	Distribuição Kolmogorov-Smirnov	98

Lista de Figuras

1.1	Comparação das funções de ligação logito, probito e complemento log-log.	4
1.2	Gráfico de probabilidade da função transformação box-cox para $\lambda = -0.5, 0, 0.5$.	5
1.3	Gráfico de probabilidade da função probito-assimétrico onde a curva de cima para baixo corresponde a $\lambda = 1, 2, 0, -1, -2$.	6
4.1	Viés encontrado na estimativa dos parâmetros das funções de ligações usando inferência bayesiana sem o uso de variáveis auxiliares.	57
4.2	Viés encontrado na estimativa dos parâmetros das funções de ligações usando inferência bayesiana com o uso de variáveis auxiliares.	58
5.1	Estudo do resíduo usando a função probito-assimétrico na inferência clássica.	61
5.2	Estudo do resíduo usando a função probito-assimétrico na inferência bayesiana.	62
5.3	Estudos dos resíduos utilizando as funções logito e probito nas inferências clássica e bayesiana, respectivamente.	67

Lista de Tabelas

4.1	Percentual de cobertura utilizando o método de máxima verossimilhança.	45
4.2	Percentual de escolha utilizando os critérios AIC e BIC.	46
4.3	Resultados das estimativas usando a função de ligação probito, utilizando os métodos com e sem o uso de variáveis auxiliares.	48
4.4	Resultados das estimativas utilizando a função de ligação logito, usando os métodos sem e com o uso variáveis auxiliares.	49
4.5	Resultados das estimativas utilizando as funções de ligações assimétricas, complemento log-log, transformação box-cox e probito-assimétrico, usando o método sem o uso variáveis auxiliares.	50
4.6	Resultados obtidos com a estimação da função de ligação probito-assimétrico, utilizando o método com o uso de variáveis auxiliares.	51
4.7	Critério de precisão para a função probito, utilizando o método com e sem o uso de variáveis auxiliares.	52
4.8	Resultados obtidos com os critérios de precisões para a função logito, usando os métodos sem e com o uso de variáveis auxiliares.	53
4.9	Resultados obtidos usando os critérios de precisões para as funções: complemento log-log, transformação box-cox e Probitto-assimétrico, utilizando o método sem o uso de variáveis auxiliares.	54
4.10	Resultados obtidos utilizando os critérios de precisões para a função de ligação probito-assimétrico, usando método com o uso de variáveis auxiliares.	55
5.1	Resultado da análise dos dados do Ibovespa utilizando inferência clássica.	60
5.2	Resultado da análise dos dados do Ibovespa utilizando inferência bayesiana.	60
5.3	Análise de dados educacionais usando inferência clássica.	63
5.4	Análise de dados educacionais usando inferência clássica, sem a variável uso de internet.	64

5.5	Resultados da análise dos dados educacionais usando inferência bayesiana com o uso de variáveis auxiliares.	65
5.6	Resultados dos cálculos da razão de chances nas abordagens clássica e bayesiana.	66

Introdução

Nelder e Wedderburn (1972) propuseram uma teoria unificadora de vários modelos estatístico e atribuíram o nome de Modelos Lineares Generalizados (MLG), que são uma extensão dos modelos lineares clássicos.

Para uma amostra aleatória de n observações, (Y_i, \mathbf{x}_i) , em que y_i é o i -ésimo elemento do vetor da variável resposta e \mathbf{x}_i é um vetor coluna formado pela i -ésima observação das p variáveis explicativas, ligação bi-paramétrica que abrange os modelos logito, probito e algumas ligações assimétricas como casos limite. 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 caso particulares. Chen et al. (1999) definiram uma nova classe de funções de ligações assimétricas para modelos de dados binários, denominado *skew probit* (probito-assimétrico), e tem como caso particular o modelo probito. Bazán et al. (2006) utilizaram uma outra função de ligação, também chamado de probito assimétrico, baseada na distribuição normal-assimétrica de Azzalini (1985).

Nesta dissertação trabalhamos com as funções de ligações probito, logito, complemento log-log, transformação box-cox e probito-assimétrico nas abordagens clássica e bayesiana. Dentro da inferência bayesiana vamos comparar duas metodologias, com e sem o uso de variáveis auxiliares. No uso de variáveis auxiliares, apresentamos e

implementamos a proposta feitas por Holmes e Held (2006) para as ligações probito e logito e na ligação probito-assimétrico trabalhamos com a ligação proposta por Chen et al. (1999) e implementamos três diferentes algoritmos de simulações proposto por Farias (2007).

1.1 Regressão Binária

Modelos para dados binários são aqueles em que a variável de interesse, habitualmente denominada por variável resposta, admite apenas dois valores. Geralmente atribui 1 para a ocorrência do evento de interesse (“sucesso”) e 0 para a ocorrência do evento complementar (“fracasso”). Essas variáveis estão comumente associadas a outras variáveis, que podem ser contínuas, discretas ou categorizadas. Consideramos que a probabilidade de sucesso possa ser explicada por estas outras variáveis, denominadas variáveis explicativas ou covariáveis.

Considere $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ o vetor formado por variáveis respostas que possuem distribuição Bernoulli, sendo Y_1, \dots, Y_n variáveis aleatórias independentes. Tal que $P(Y_i = 1) = \mu_i$ e $P(Y_i = 0) = 1 - \mu_i$. Considere também $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ o vetor formado por variáveis explicativas associadas a Y_i , previamente fixadas, onde x_{i1} pode ser igual a 1, denominado intercepto.

A distribuição Bernoulli é dada por

$$P(Y_i = y_i) = \mu_i^{y_i} (1 - \mu_i)^{1-y_i}, \quad y_i \in \{0, 1\}.$$

Colocando a na família exponencial, temos que

$$P(Y_i = y_i) = \exp \left\{ y_i \log \left(\frac{\mu_i}{1 - \mu_i} \right) + \log(1 - \mu_i) \right\}.$$

Assim, o parâmetro canônico θ_i é dado por: $\theta_i = \log \left(\frac{\mu_i}{1 - \mu_i} \right)$, $b(\theta_i) = \log(1 + e^{\theta_i}) = -\log(1 - \mu_i)$, sendo que $E(Y_i) = b'(\theta_i) = \mu_i$.

Na modelagem de dados binários assume-se que

$$\mu_i = F(\eta_i) = F(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad i = 1, \dots, n$$

sendo $F(\cdot)$ é uma função de distribuição acumulada e $F^{-1}(\mu_i)$ é chamada de função de ligação na teoria de modelo linear generalizados (MLG), e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é um vetor de dimensão p de coeficientes regressores.

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

1.2 Funções de Ligações

A escolha da função de ligação deve ser feita com a finalidade de facilitar a interpretação do modelo e sempre deve estar de acordo com a distribuição proposta para os dados. Para a distribuição Bernoulli temos interesse em encontrar uma função de ligação que transforma o intervalo $]0, 1[$ na reta dos reais. Para isso, utilizamos funções de ligações simétricas, probito e logito, e ligações assimétricas, complemento log-log, transformação box-cox e probito-assimétrico. Na ligação complemento log-log, não existe um parâmetro que modele assimetria. Diferente das ligações box-cox e probito-assimétrico que tem um parâmetro adicional.

1.2.1 Logito

Considere a função de distribuição acumulada da distribuição logística padrão dada por

$$F(\mathbf{x}_i^\top \boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}.$$

Neste caso $F^{-1}(\mu_i)$ é função de ligação Logito, $g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$.

1.2.2 Probit

Seja $F(\cdot)$ a função de distribuição acumulada da normal padrão. A regressão probito é dada por: $F(\mathbf{x}_i^\top \boldsymbol{\beta}) = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})$ e $F^{-1}(\mu_i)$ é a função de ligação Probit, $g(\mu_i) = \Phi^{-1}(\mu_i)$.

1.2.3 Complemento log-log

Considere $F(\cdot)$ a função de distribuição acumulada da distribuição Gumbel, isto é

$$F(\mathbf{x}_i^\top \boldsymbol{\beta}) = 1 - \exp(-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})).$$

Neste caso, $F^{-1}(\mu_i)$ é a função de ligação complemento log-log, $g(\mu_i) = \log(-\log(1 - \mu_i))$.

Na Figura 1.1 mostramos o comportamento de μ em função de η considerando as três funções, logito, probito e complemento log-log. A ligação complemento log-log é capaz de modelar melhor probabilidades próximas a 1, mas comporta-se como logito para probabilidades baixas. A ligação probito é próxima da logito, mas aproxima-se dos extremos ligeiramente mais rápido.

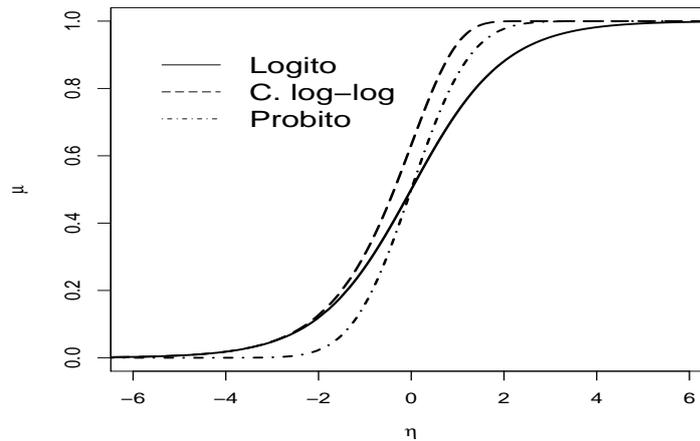


Figura 1.1: Comparação das funções de ligação logito, probito e complemento log-log.

1.2.4 Transformação box-cox

Segundo Guerrero e Johnson (1982) a função de transformação de box-cox é dada por

$$g(\mu_i) = \begin{cases} \frac{\left(\frac{\mu_i}{1-\mu_i}\right)^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0, \\ \log\left(\frac{\mu_i}{1-\mu_i}\right) & \text{se } \lambda = 0. \end{cases}$$

Quando $\lambda = 0$ o modelo logito é um caso particular da função box-cox.

Na Figura 1.2 mostramos o comportamento de μ em função de η considerando a função de transformação box-cox para valores de $\lambda = -0.5, 0, 0.5$.

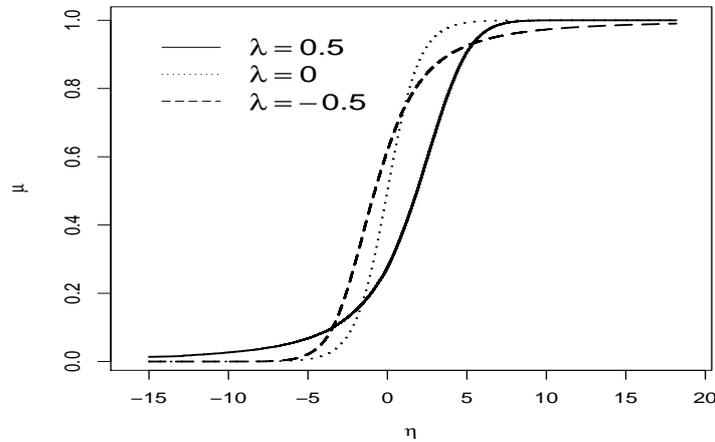


Figura 1.2: Gráfico de probabilidade da função transformação box-cox para $\lambda = -0.5, 0, 0.5$.

1.2.5 Probito-assimétrico

Considere $F_\lambda(\cdot)$ a função de distribuição acumulada da normal-assimétrica definida por Azzalini (1985). O modelo probito-assimétrico dado por

$$F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta}) = \Phi_{SN}(\mathbf{x}_i^\top \boldsymbol{\beta}; \mu, \sigma^2, \lambda), \quad (1.1)$$

sendo $\Phi_{SN}(\cdot; \mu, \sigma^2, \lambda)$ a fda da distribuição normal-assimétrica, com o parâmetros de localização μ , de escala σ^2 e de assimetria λ .

A função de ligação $F_\lambda^{-1}(\cdot)$, é a inversa da distribuição normal-assimétrica (Apêndice C), logo é a função de ligação probito-assimétrico. Para $\lambda = 0$ tem-se modelo probito.

Na Figura 1.3 mostramos o comportamento de μ em função de η considerando a função probito-assimétrico para valores de $\lambda = 1, 2, 0, -1, -2$.

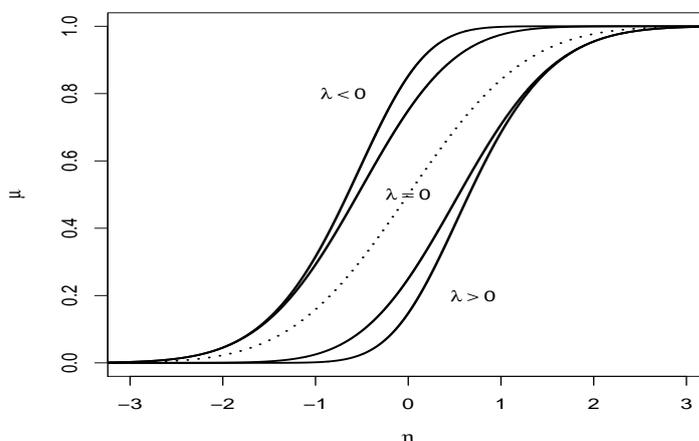


Figura 1.3: Gráfico de probabilidade da função probito-assimétrico onde a curva de cima para baixo corresponde a $\lambda = 1, 2, 0, -1, -2$.

1.3 Objetivos e Organização da Dissertação

O objetivo deste trabalho é apresentar o modelo de regressão binária no contexto clássico e bayesiano. Para isso, este trabalho está dividido em duas partes. Na primeira parte apresentamos inferência clássica usando o método de máxima verossimilhança para estimar os parâmetros. Na segunda parte apresentamos a inferência bayesiana com e sem o uso de variáveis auxiliares. Para o estudo com o uso das variáveis auxiliares utilizamos as funções simétricas e assimétricas. Para as funções simétricas, probito e logito, apresentamos e implementamos os algoritmos de simulação propostos por Holmes e Held (2006). Na função assimétrica, probito-assimétrico, propostos por Chen et al. (1999), implementamos a proposta feita por Farias (2007).

Essa dissertação está dividida em seis capítulos e quatro apêndices. No segundo capítulo, apresentamos a metodologia clássica, em que, utilizamos o método de máxima verossimilhança para estimar os parâmetros. Além disso, apresentamos as metodologias que foram utilizadas para análise dos resíduos.

No terceiro capítulo, apresentamos a metodologia bayesiana. Na primeira parte, apresentamos a inferência bayesiana sem o uso de variáveis auxiliares, para todas as funções de ligações apresentadas no primeiro capítulo. Na segunda parte, apresentamos a inferência bayesiana com o uso de variáveis auxiliares, nos modelos de regressão simétricos, probito e logito, e o modelo de regressão assimétrico, probito-assimétrico. Na função probito-assimétrico utilizamos dois conjuntos de variáveis auxiliares, com três algoritmos para obtenção da distribuição *a posteriori* conjunta dos parâmetros. Por

fim, apresentamos o método que utilizamos para fazer o estudo de pontos influentes e os critérios de eficiência para comparar os algoritmos utilizados na inferência bayesiana.

No quarto capítulo, apresentamos os resultados obtidos com os estudos de simulações. No quinto capítulo, realizamos uma análise de um conjunto de dados reais. Finalmente, no sexto capítulo, apresentamos as considerações finais e perspectivas futuras.

No apêndice A, apresentamos os pseudo-códigos do modelo bayesiano com uso de variáveis auxiliares para implementação computacional dos algoritmos. No apêndice B, mostramos como gerar valores da distribuição *a posteriori* da variável de mistura no modelo logístico. No apêndice C, encontram-se as demonstrações relativas à distribuição *a posteriori* da função probito-assimétrico. No apêndice D, apresentamos algumas funções de distribuições de probabilidade utilizadas ao longo do texto.

Abordagem Clássica

Nesta Seção apresentamos o método utilizado para estimar os parâmetros, construção dos intervalos de confiança, métodos diagnósticos, critérios de seleção das funções de ligações e, por fim, as propriedades dos estimadores.

2.1 Estimação dos Parâmetros

Seja a função de probabilidade da distribuição Bernoulli dada por

$$P(Y_i = y_i) = \mu_i^{y_i} (1 - \mu_i)^{1-y_i}, \quad y_i \in \{0, 1\}. \quad (2.1)$$

Considere uma amostra com n observações independentes, dada por (y_i, \mathbf{x}_i) , $i = 1, \dots, n$, em que y_i é a i -ésima observação da variável resposta e \mathbf{x}_i é um vetor coluna formado pela i -ésima observação das p variáveis explicativas e uma função de ligação genérica $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$, com $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$. Para estimar os parâmetros β_1, \dots, β_p utilizaremos o método de máxima verossimilhança.

Assumindo a independência entre as observações, a verossimilhança é obtida por

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n P(Y_i = y_i) = \prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1-y_i}$$

A log-verossimilhança é definida por

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \log \left[\prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1-y_i} \right] \\
&= \sum_{i=1}^n \{y_i \log(\mu_i) + (1 - y_i) \log(1 - \mu_i)\} \\
&= \sum_{i=1}^n \{y_i \log(\mu_i) + \log(1 - \mu_i) - y_i \log(1 - \mu_i)\} \\
&= \sum_{i=1}^n \left\{ y_i \log \left(\frac{\mu_i}{1 - \mu_i} \right) + \log(1 - \mu_i) \right\} \tag{2.2}
\end{aligned}$$

Utilizando

$$\theta_i = \log \left(\frac{\mu_i}{1 - \mu_i} \right) \Rightarrow \mu_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}},$$

podemos simplificar (2.2), que torna-se

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \{y_i \theta_i - b(\theta_i)\}, \tag{2.3}$$

sendo

$$b(\theta_i) = -\log(1 - \mu_i) = -\log \left(1 - \frac{e^{\theta_i}}{1 + e^{\theta_i}} \right) = \log(1 + e^{\theta_i}).$$

Usando as relações

$$\eta_i = \mathbf{x}'_i \boldsymbol{\beta}; \quad \mu_i = g^{-1}(\eta_i); \quad \theta_i = \log \left(\frac{\mu_i}{1 - \mu_i} \right),$$

obtemos as funções Score para uma função de ligação qualquer, que são

$$\begin{aligned}
U_{\beta_r} &= \frac{\partial \ell}{\partial \beta_r} = \sum_{i=1}^n \frac{d\ell}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_r}, \quad r = 1, \dots, p \\
U_{\beta_r} &= \sum_{i=1}^n \{y_i - b'(\theta_i)\} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} x_{ir}.
\end{aligned}$$

Usando $\mu_i = b'(\theta_i)$ e $V_i = \frac{d\mu_i}{d\theta_i}$ podemos escrever

$$U_{\beta_r} = \sum_{i=1}^n \{y_i - \mu_i\} \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir}$$

Como

$$\frac{d\theta_i}{d\mu_i} = \frac{1 - \mu_i (1 - \mu_i) + \mu_i}{\mu_i (1 - \mu_i)^2} = \frac{1}{\mu_i(1 - \mu_i)},$$

então

$$\frac{d\mu_i}{d\theta_i} = V_i = \mu_i(1 - \mu_i).$$

Como

$$\eta_i = g(\mu_i) \Rightarrow \mu_i = g^{-1}(\eta_i),$$

então

$$\frac{d\mu_i}{d\eta_i} = \frac{dg^{-1}(\eta_i)}{d\eta_i}.$$

Logo,

$$U_{\beta_r} = \sum_{i=1}^n \{y_i - \mu_i\} \frac{1}{V_i} \frac{dg^{-1}(\eta_i)}{d\eta_i} x_{ir}, \quad (2.4)$$

para $r = 1, \dots, p$.

O estimador de máxima verossimilhança $\hat{\beta}$ do vetor de parâmetros β é obtido igualando U_r a zero para $r = 1, \dots, p$. Usaremos o método de Newton-Rapshon, já que, no nosso caso, a função log-verossimilhança é não linear nos parâmetros.

Usaremos o método escore de Fisher dado por

$$\beta^{(m+1)} = \beta^{(m)} + (\mathbf{K}^{(m)})^{-1} \mathbf{U}^{(m)} \quad (2.5)$$

sendo que \mathbf{K} tem elementos típicos dados por

$$k_{r,s} = -E \left[\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s} \right] = E(U_{\beta_r} U_{\beta_s})$$

que é a matriz de covariâncias dos U'_r 's.

Para o caso da distribuição Bernoulli, as derivadas em relação aos β 's é dada por

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s} = - \sum_{i=1}^n x_{ir} x_{is} \mu_i (1 - \mu_i) \quad (2.6)$$

para $r, s = 1, 2, \dots, p$.

Multiplicando-se ambos os membros de (2.5) por $\mathbf{K}^{(m)}$, tem-se

$$\mathbf{K}^{(m)} \beta^{(m+1)} = \mathbf{K}^{(m)} \beta^{(m)} + \mathbf{U}^{(m)} \quad (2.7)$$

O elemento típico k_{rs} de \mathbf{K} é obtido de (2.4) como

$$k_{r,s} = E(U_{\beta_r} U_{\beta_s}) = \sum_{i=1}^n E(Y_i - \mu_i)^2 \frac{1}{V_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ir} x_{is}$$

ou

$$k_{r,s} = E(U_{\beta_r} U_{\beta_s}) = \phi^{-1} \sum_{i=1}^n w_i x_{ir} x_{is},$$

sendo $w_i = V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right)^2$ denominado função peso. Logo, a matriz de informação de Fisher para $\boldsymbol{\beta}$ tem a forma

$$\mathbf{K} = \mathbf{X}^\top \mathbf{W} \mathbf{X},$$

sendo $\mathbf{W} = \text{diag} \{w_1, \dots, w_n\}$ uma matriz diagonal de pesos. No caso das funções de ligação canônicas tem-se $w_i = V_i$, pois $V_i = V(\mu_i) = d\mu_i/d\eta_i$. A ligação canônica para a distribuição Bernoulli é dada por $\eta_i = \log \left\{ \frac{\mu_i}{1 - \mu_i} \right\}$. Fazendo os cálculos necessários, podemos ver que a matriz \mathbf{W} é dada por

$$\mathbf{W} = \begin{bmatrix} \hat{\mu}_1(1 - \hat{\mu}_1) & 0 & \cdots & 0 \\ 0 & \hat{\mu}_2(1 - \hat{\mu}_2) & \cdots & 0 \\ \vdots & 0 & \ddots & \cdots \\ 0 & \cdots & 0 & \hat{\mu}_n(1 - \hat{\mu}_n) \end{bmatrix}$$

O vetor escore $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$ com componentes em (2.4) pode, então ser escrito na forma

$$\mathbf{U} = \mathbf{X}^\top \mathbf{W} \mathbf{G}(\mathbf{y} - \boldsymbol{\mu})$$

com $\mathbf{G} = \text{diag} \{d\eta_1/d\mu_1, \dots, d\eta_n/d\mu_n\} = \text{diag} \{g'(\mu_1), \dots, g'(\mu_n)\}$. Substituindo \mathbf{K} e \mathbf{U} em (2.7), temos que

$$\mathbf{X}^\top \mathbf{K}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$$

Colocando $\mathbf{X}^\top \mathbf{W}^{(m)}$ em evidência obtemos

$$\mathbf{X}^\top \mathbf{K}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^\top \mathbf{W}^{(m)} \left[\boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \right]$$

Então $\boldsymbol{\beta}$ é dado por

$$\boldsymbol{\beta}^{(m+1)} = \left(\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (2.8)$$

em que, $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})$.

2.1.1 Estimação para as Funções de Ligações Assimétricas

Para as funções de ligações transformação box-cox e probito-assimétrico podemos considerar λ um valor fixo o qual deve ser proposto visando a facilidade na interpretação. Alternativamente, pode-se estimar λ por máxima verossimilhança. A função log-verossimilhança como função de λ é dada por

$$\ell(\lambda, \boldsymbol{\beta}) = \sum_{i=1}^n \left\{ y_i \log \left(\frac{\mu_i}{1 - \mu_i} \right) + \log(1 - \mu_i) \right\}. \quad (2.9)$$

Utilizando a função transformação box-cox obtemos

$$\eta_i = g(\mu_i) = \frac{\left(\frac{\mu_i}{1 - \mu_i} \right)^\lambda - 1}{\lambda}; \quad \lambda \neq 0,$$

o que implica

$$\mu_i = g^{-1}(\eta_i) = \frac{1}{1 + (1 + \eta_i \lambda)^{-1/\lambda}}; \quad \lambda \neq 0.$$

Utilizando a função probito-assimétrico temos

$$\eta_i = g(\mu_i) = \Phi_{SN}^{-1}(\mu_i; \mu, \sigma^2, \lambda),$$

a partir da qual obtemos

$$\mu_i = g^{-1}(\eta_i) = \Phi_{SN}(\eta_i; \mu, \sigma^2, \lambda),$$

sendo $\Phi_{SN}(\cdot; \mu, \sigma^2, \lambda)$ a fda da distribuição normal-assimétrica, com os parâmetros de localização μ , de escala σ^2 e de assimetria λ .

A maximização da equação (2.9) em relação a λ apresenta problemas computacionais e deve ser feita em duas partes. Fixa-se $\lambda = \lambda^*$ e maximiza-se $\ell(\lambda^*, \boldsymbol{\beta})$ em relação aos demais parâmetros. Com o valor de $\boldsymbol{\beta}$ estimado encontra-se o valor de λ que maximiza a função de verossimilhança.

Denotando-se por $\hat{\ell}$ o valor da função log-verossimilhança nos pontos λ e $\boldsymbol{\beta}$.

É informativo traçar o gráfico de $\widehat{\ell}(\lambda, \boldsymbol{\beta})$ versus λ para um certo conjunto de valores deste parâmetro, por exemplo, os inteiros de -3 a 3. A estimativa de λ corresponderá ao ponto de maior $\widehat{\ell}(\lambda, \boldsymbol{\beta})$.

2.2 Intervalo de Confiança

Geralmente não é possível encontrar distribuições exatas para os estimadores, assim sendo, trabalha-se com resultados assintóticos, considerando que o modelo escolhido irá satisfazer as condições de regularidade. Cox e Hinkley (1986) demonstraram que, em problemas regulares, a função *Score* $\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ tem valor esperado igual a zero e a estrutura de covariância é igual a matriz de informação de Fisher, dado por

$$\mathbf{K} = \mathbf{I}(\boldsymbol{\beta}) = \mathbf{E} \left[\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_r \partial \beta_s} \right] = \mathbf{X}^\top \mathbf{W} \mathbf{X} \quad (2.10)$$

Assim, a distribuição assintótica dos $\boldsymbol{\beta}'s$ é dada por

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{K} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_p^2 \quad (2.11)$$

ou, de forma equivalente

$$\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{K}^{-1}). \quad (2.12)$$

Os erros padrão dos EMV $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ são iguais às raízes quadradas dos elementos de diagonal principal de $\widehat{\mathbf{K}}^{-1}$. Usando a notação $\mathbf{K}^{-1} = \{k^{r,s}\}$ para a inversa da matriz de informação. Então, ao nível de confiança de 95%, o intervalo de confiança para os parâmetros $\boldsymbol{\beta}'s$ é

$$\widehat{\beta}_r \pm 1,96 \sqrt{\widehat{k}^{r,r}},$$

em que $\widehat{k}^{r,r} = \widehat{Var}(\widehat{\beta}_r)$ é o valor de $\widehat{k}^{r,r}$ em $\widehat{\boldsymbol{\beta}}$.

A correlação estimada entre as estimativas $\widehat{\beta}_r$ e $\widehat{\beta}_s$ é

$$\widehat{Corr}(\widehat{\beta}_r, \widehat{\beta}_s) = \frac{\widehat{k}^{r,s}}{\sqrt{\widehat{k}^{r,r} \widehat{k}^{s,s}}},$$

sendo obtida diretamente da inversa da matriz de informação \mathbf{K} avaliada em $\widehat{\boldsymbol{\beta}}$.

2.3 Testes de Hipóteses

Existem três estatísticas assintoticamente equivalentes utilizadas para testar hipóteses relativas aos parâmetros β 's. Primeiramente temos o teste de razão de verossimilhança, indicado para as hipóteses que envolvem vários coeficientes. Quando estamos considerando um único coeficiente β_r , é aconselhável usar o teste Wald. Por último, temos o teste escore que tem sido muito usado em Bioestatística.

Supomos que temos interesse em testar as hipóteses

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0 \end{cases}$$

Sob a hipótese nula H_0 , as três estatísticas convergem para uma variável aleatória com distribuição χ_p^2 .

As três estatísticas para testar a hipótese nula H_0 são apresentadas a seguir.

2.3.1 Teste de razão de verossimilhanças

O teste de razão de verossimilhança é definido por

$$\xi_{RV} = 2 \left\{ \ell(\hat{\beta}) - \ell(\beta_0) \right\},$$

em que $\ell(\hat{\beta})$ e $\ell(\beta_0)$ são os valores da verossimilhança anulada em $\hat{\beta}$ e β_0 , respectivamente.

2.3.2 Teste Wald

A estatística de Wald (1943) é baseada na distribuição normal assintótica de $\hat{\beta}$. Ela é definida por

$$\xi_W = (\hat{\beta} - \beta_0)^\top \hat{\mathbf{K}} (\hat{\beta} - \beta_0),$$

em que $\hat{\mathbf{K}}$ é a matriz de informação avaliada em $\hat{\beta}$.

2.3.3 Teste Escore

A estatística escore Rao et al. (1973) é obtida através da função escore, dada por

$$\xi_{SR} = \mathbf{U}(\beta_0)^\top \mathbf{K}_0^{-1} \mathbf{U}(\beta_0),$$

em que $\mathbf{U}(\boldsymbol{\beta}_0)$ é o vetor escore, \mathbf{K}_0 é a matriz de informação avaliada em $\hat{\boldsymbol{\beta}}$.

2.4 Função Desvio

A função desvio é calculada com a finalidade medir a discrepância entre o modelo saturado e o modelo proposto. O modelo saturado é aquele que contém tantos parâmetros quantas observações amostrais. Um valor pequeno da função desvio indica que, para um menor número de parâmetros, obtém-se um ajuste tão bom quanto o ajuste com o modelo saturado.

Nelder e Wedderburn (1972) propuseram, como a medida de discrepância, o *deviance*, dada por

$$D_p = 2(\hat{\ell}_n - \hat{\ell}_p)$$

sendo $\hat{\ell}_n$ e $\hat{\ell}_p$ os máximos da log-verossimilhança para os modelos saturado e proposto, respectivamente.

Do logaritmo da função de verossimilhança, obtém-se

$$\hat{\ell}_n = \sum_{i=1}^n [y_i \tilde{\theta}_i - b(\tilde{\theta}_i)] + \sum_{i=1}^n c(y_i)$$

e

$$\hat{\ell}_p = \sum_{i=1}^n [y_i \hat{\theta}_i - b(\hat{\theta}_i)] + \sum_{i=1}^n c(y_i),$$

sendo $\tilde{\theta}_i = q(y_i)$ e $\hat{\theta}_i = q(\hat{\mu}_i)$ o EMV do parâmetro canônico sob os modelos saturado e proposto, respectivamente. Então,

$$D_p = 2 \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)].$$

Considerando o $Y_i \sim \text{Bernoulli}(\mu_i)$, $i = 1, \dots, n$, obtemos

$$\tilde{\theta}_i = \log\left(\frac{y_i}{1-y_i}\right) \quad ; \quad b(\tilde{\theta}_i) = \log\left(\frac{1}{1-y_i}\right)$$

e

$$\hat{\theta}_i = \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) \quad ; \quad b(\hat{\theta}_i) = \log\left(\frac{1}{1-\hat{\mu}_i}\right)$$

Logo, o desvio assume a seguinte forma

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (1 - y_i) \log \left(\frac{1 - y_i}{1 - \hat{\mu}_i} \right) \right].$$

Quando $y_i = 0$ ou $y_i = 1$, o i -ésimo termo de $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ vale $-2 \log(1 - \hat{\mu}_i)$ ou $-2 \log \hat{\mu}_i$, respectivamente. O desvio é sempre positiva e quanto menor, melhor é o ajuste do modelo.

2.5 Diagnósticos Clássicos de Regressão

Nos MLGs, segundo Cordeiro (2006), os resíduos são usados para explorar a adequação do modelo ajustado em relação à escolha da função de variância, da função de ligação e de termos no preditor linear. Além disso, eles também são úteis na identificação de pontos aberrantes, que poderão ser influentes ou não. Segundo Cox e Snell (1968), os resíduos devem expressar a discrepância (distância) entre a observação y_i e o seu valor ajustado $\hat{\mu}_i$.

2.5.1 Resíduos de Pearson

O resíduo de *Pearson* auxilia na classificação de uma observação, que pode ser considerado como *outliers*.

Na regressão logística, o resíduo de *Pearson* é definido por

$$r(y_j, \hat{\mu}_j) = \frac{y_j - \hat{\mu}_j}{\sqrt{\hat{\mu}_j(1 - \hat{\mu}_j)}} \quad (2.13)$$

Assim, a estatística qui-quadrado de Pearson é dada por

$$\chi^2 = \sum_{i=1}^j r(y_i, \hat{\mu}_j)^2.$$

Esta estatística possui distribuição assintótica χ_{n-p}^2 . A desvantagem de resíduo de Pearson é que sua distribuição é, geralmente, bastante assimétrica para modelos não-normais.

2.5.2 Resíduos de Anscombe

Anscombe (1953) apresenta uma definição geral de resíduo, através de uma transformação $\psi(y_i)$ da observação y_i , escolhida visando tornar a sua distribuição o mais próxima da distribuição normal. Barndorff-Nielsen (1978) demonstra que, para os MLG, $\psi(\cdot)$ é dada por $\psi(\mu) = \int V^{-1/3} d\mu$. Como $\widehat{V}^{1/2}(\widehat{\mu}_i)\psi'(\widehat{\mu}_i)$ é a aproximação de primeira ordem do desvio de $\psi(\mu)$, o resíduo de Anscombe, visando a normalidade e à estabilidade da variância, é dado por

$$A_i = \frac{\psi(y_i) - \psi(\widehat{\mu}_i)}{\widehat{V}^{1/2}(\widehat{\mu}_i)\psi'(\widehat{\mu}_i)} \quad (2.14)$$

em que, $\psi(\cdot)$ é a função transformação utilizada pra normalizar a distribuição Y .

Para o modelo Bernoulli com média μ , a Equação (2.14) se reduz $A_i = [\psi(y_i) - \psi(\widehat{\mu}_i)]/[\widehat{\mu}_i(1 - \widehat{\mu}_i)]^{1/6}$, em que $\psi(\mu) = \int [\mu(1 - \mu)]^{-1/3} d\mu$.

2.5.3 Resíduos de Pearson Studentizados

Os resíduos studentizados são dados por

$$rp_j = \frac{y_j - \widehat{\mu}_j}{\sqrt{V(\widehat{\mu}_j)(1 - \widehat{h}_{jj})}}, \quad (2.15)$$

sendo h_{jj} é o j -ésimo elemento da diagonal da matriz \mathbf{H} dada por

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{1/2}.$$

em que, \mathbf{W} é uma matriz diagonal com elementos $\widehat{\mu}_j(1 - \widehat{\mu}_j)$, \mathbf{X} é um matriz $n \times p$.

2.5.4 Componentes do Desvio

Componente do desvio é o mais utilizado em MLG's, e podemos defini-lo como a raiz quadrada da diferença entre as log-verossimilhanças dos modelos saturado e proposto para cada uma das observações, com sinal de $y_i - \widehat{\mu}_i$, ou seja,

$$r_j = \text{sin}al(y_j - \widehat{\mu}_j) \sqrt{2(\widehat{\ell}_n - \widehat{\ell}_p)} \quad (2.16)$$

sendo $\widehat{\ell}_n$ e $\widehat{\ell}_p$ os máximos do logaritmo da função de verossimilhança para o modelo saturado e proposto.

Como podemos observar, o resíduo r_j representa uma distância da observação y_i ao valor ajustado $\hat{\mu}_i$, medida na escala do logaritmo da função de verossimilhança.

2.5.5 Componentes do Desvio Estudentizados

Os resíduos do desvio studentizados é dado por

$$r'_j = \frac{r_j}{\sqrt{1 - \hat{h}_{jj}}} \quad (2.17)$$

sendo \hat{h}_{jj} é o j -ésimo elemento da digonal da matriz \mathbf{H} .

Williams (1984) verificou que a distribuição r_j está mais próxima da normal. Assim, para grandes amostras, são considerados marginalmente aberrantes pontos tais que

$$|r'_j| > 2.$$

2.5.6 Distância de Cook

A distância de Cook, representada por D_i , mede o afastamento do vetor de estimativas dos coeficientes da regressão provocado pela retirada da i -ésima observação. A expressão usa como estimativa da variância residual aquela obtida com todas as n observações, ou ainda, considera o resíduo studentizado internamente. A estatística é dado por

$$D_i \cong \left[\frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} \right] r p_i^2.$$

2.6 Critério de Seleção de Modelos

Ao selecionarmos modelos é preciso ter em mente que não existem modelos verdadeiros, há apenas modelos aproximados da realidade que, causam perda de informações. Deste modo, é necessário fazer a seleção do melhor modelo, entre aqueles que estão bem ajustados, para explicar o fenômeno sob estudo.

Busca-se o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Nesta linha, diversos critérios para seleção de modelos são apresentados na literatura (ver Bozdogan (1987); Wolfinger (1993); Littell et al. (2006)). Dentre os critérios para seleção de modelos, os critérios baseados no máximo da verossimilhança (MFV) são os mais utilizados, com

maior ênfase o Teste da Razão de Verossimilhança (TRV), o Critério de Informação de Akaike (AIC) e o Critério Bayesiano de Schwarz (BIC).

2.6.1 Critério de Akaike - AIC

Um dos critérios de informação que tem sido muito usado em seleção de modelos é o critério Akaike (AIC), Akaike (1974). Este critério admite a existência de um modelo “real”, desconhecido, que descreve os dados, e tenta escolher entre um grupo de modelos avaliados, o que minimiza a divergência de Kullback-Leibler (K-L). O valor de K-L para um modelo f com parâmetros β , em relação ao modelo “real” representado por g é

$$\text{KL}(f; g) = \int g(y) \log \left(\frac{g(y)}{f(y|\beta)} \right) dy.$$

Esta divergência está relacionada à informação perdida por se usar um modelo aproximado e não o “real”. A estimativa do AIC para um determinado modelo é dada por

$$\text{AIC} = -2 \sum_{i=1}^n \log L(\hat{\beta}; y_i) + 2p,$$

em que, y_i é o i -ésimo valor da variável resposta, $\hat{\beta}$ é a estimativa de y_i e p é o número de parâmetros. Entre vários modelos candidatos, deve ser escolhido aquele que apresenta o menor valor de AIC.

2.6.2 Critério de Informação Bayesiano - BIC

O Critério de Informação Bayesiano (BIC), proposto por Schwarz et al. (1978) é dado por

$$\text{BIC} = -2 \sum_{i=1}^n \log L(\hat{\beta}, y_i) + p \log(n),$$

sendo n é o tamanho da amostra.

Baseia-se, em parte, na função de probabilidade e está intimamente relacionado com o critério de informação de Akaike (AIC). Ao montar modelos, é possível aumentar a probabilidade adicionando parâmetros, mas fazer isso pode resultar em superajuste.

Os critérios AIC e BIC tentam resolver este problema através da introdução de um termo de penalidade para o número de parâmetros do modelo, este termo é mais alto no BIC do que no AIC. O modelo com menor valor é considerado o de melhor ajuste.

2.7 Propriedades dos Estimadores

Definição 2.1. Seja uma amostra aleatória de uma distribuição parametrizada por β . O erro quadrático médio de um estimador $\hat{\beta}$ de β é definido como

$$\text{EQM}(\hat{\beta}) = \text{E}(\hat{\beta} - \beta)^2.$$

Podemos reescrevendo esta última expressão da seguinte maneira

$$\begin{aligned} \text{EQM}(\hat{\beta}) &= \text{E}[(\hat{\beta} - \text{E}(\hat{\beta})) + (\text{E}(\hat{\beta}) - \beta)]^2 \\ &= \text{E}[\hat{\beta} - \text{E}(\hat{\beta})]^2 + [\text{E}(\hat{\beta}) - \beta]^2 \\ &= \text{Var}(\hat{\beta}) + [\text{E}(\hat{\beta}) - \beta]^2. \end{aligned}$$

sendo o termo $\text{E}(\hat{\beta}) - \beta$ é chamado *vício* ou *viés* do estimador e denotado por $B(\beta)$.

Assim, o erro quadrático médio é definido como a variância do estimador mais o quadrado do seu viés. Um caso particular ocorre quando $B(\beta) = 0$, ou equivalentemente $\text{E}(\hat{\beta}) = \beta$. Neste caso, $\hat{\beta}$ é um estimador não viesado (ENV) para β . Logo, pela definição 2.1 temos que $\text{EQM}(\hat{\beta}) = \text{Var}(\hat{\beta})$. Se $\text{E}(\hat{\beta}) \neq \beta$ então o estimador $\hat{\beta}$ é dito *viesado* ou *viciado*.

Quando aumentamos o tamanho da amostra, a esperança do estimador se aproxima do verdadeiro valor de β , $\lim_{n \rightarrow \infty} \text{E}(\hat{\beta}) = \beta$, neste caso $\hat{\beta}$ é um estimador *assintoticamente não viesado* para β .

2.8 Considerações Finais

Concluimos, a abordagem clássica tem sido amplamente utilizada no ajuste de modelos de regressão binária, no entanto, dependendo do tipo de função de ligação adotado, esta abordagem pode esbarrar em dificuldades para encontrar estimadores de máxima verossimilhança. Além disso, realizar inferência sobre os parâmetros dos modelos com a abordagem clássica é preciso sempre ter em mente que estamos usando a teoria assintótica, válida quando o tamanho da amostra tende para o infinito e isso nem sempre é possível.

Abordagem Bayesiana

3.1 Inferência Bayesiana Sem Variáveis Auxiliares

O modelo bayesiano sem uso de variáveis auxiliares é dado por

$$\begin{aligned}y_i &\sim \text{Bernoulli}(F(\eta_i)) \\ \eta_i &= \mathbf{x}_i^\top \boldsymbol{\beta} \\ \boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}).\end{aligned}\tag{3.1}$$

Em que $y_i \in \{0, 1\}$, $i = 1, \dots, n$, 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, $\boldsymbol{\beta}$ representa um vetor coluna ($p \times 1$) de coeficientes regressores e $\pi(\cdot)$ a distribuição *a priori* de $\boldsymbol{\beta}$.

O vetor de parâmetros desconhecidos, $\boldsymbol{\beta}$, associa-se um modelo probabilístico, denominada de distribuição *a priori*. Neste trabalho consideramos priori subjetiva. Com isso, a distribuição *a posteriori* conjunta para os parâmetros, é obtida combinando essa distribuição *a priori* com a verossimilhança.

Utilizando o teorema de Bayes, escrevendo $\mathbf{y} = (y_1, \dots, y_n)^\top$, a distribuição dos parâmetros atualizada pelos dados, é a distribuição *a posteriori* dada por

$$\pi(\boldsymbol{\beta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\beta}) \prod_{i=1}^n F(\eta_i)^{y_i} [1 - F(\eta_i)]^{1-y_i}}{\int_{-\infty}^{+\infty} \prod_{i=1}^n F(\eta_i)^{y_i} [1 - F(\eta_i)]^{1-y_i} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}} \quad (3.2)$$

Como $\boldsymbol{\beta}$ é o vetor de parâmetros, temos dificuldade em calcular a integral, sendo necessário, o uso de métodos numéricos para gerar amostras aleatórias independentes dessa distribuição *a posteriori* para cada parâmetro. Existem diversos algoritmos capaz de executar essa tarefa, como Metropolis-Hastings, Hastings (1970), Amostrador de Gibbs, Geman e Geman (1984) ou o algoritmo de rejeição adaptativa. Todos fazem parte de uma família chamado Monte Carlo via cadeia de Markov (*Markov Chain Monte Carlo*: MCMC). Esses métodos são baseados na construção de cadeias de Markov cuja distribuição de interesse é a distribuição estacionária. Neste contexto, a distribuição estacionária é a distribuição *a posteriori*.

No caso da inferência bayesiana sem o uso de variáveis auxiliares utiliza-se o algoritmo de Metropolis-Hastings para gerar amostra da distribuição *a posteriori*. Considere $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^\top$, e cada componente $\boldsymbol{\beta}_i$ pode ser escalar ou um vetor e suponha que a cadeia esteja no estado $\boldsymbol{\beta}$ e um valor θ^* é gerado de uma distribuição $q(\cdot|\boldsymbol{\beta})$. O novo valor θ^* é aceito com probabilidade

$$\alpha(\boldsymbol{\beta}, \theta^*) = \min \left(1, \frac{\pi(\theta^*)q(\boldsymbol{\beta}|\theta^*)}{\pi(\boldsymbol{\beta})q(\theta^*|\boldsymbol{\beta})} \right).$$

O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos,

1. Inicializa-se o contador de iterações $t = 0$ e especifique os valores iniciais para $\boldsymbol{\beta}^{(0)} = (\boldsymbol{\beta}_1^{(0)}, \dots, \boldsymbol{\beta}_p^{(0)})^\top$.
2. Gere um novo valor θ^* da distribuição $q(\cdot|\theta)$.
3. Calcule a probabilidade de aceitação $\alpha(\boldsymbol{\beta}, \theta^*)$ e gere o valor u da variável $U \sim U(0, 1)$.
4. Se $u \leq \alpha$ então aceite o novo valor e faça $\boldsymbol{\beta}^{(t+1)} = \theta^*$, caso contrário rejeite e faça $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)}$.
5. Altere o contador de (t) para $(t + 1)$ e volte ao passo 2.

No caso das funções de ligações assimétricas, há um parâmetro adicional, λ . Neste caso, gera-se $\boldsymbol{\beta}^{(t+1)}|\lambda^{(t)}$, em seguida gera-se $\lambda^{(t+1)}|\boldsymbol{\beta}^{(t+1)}, \lambda^{(t)}$.

3.2 Inferência Bayesiana com Variáveis Auxiliares

3.2.1 Regressão Probit

Podemos representar o modelo probito utilizando uma variável auxiliar da seguinte maneira

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

Considere a distribuição *a priori*, $\pi(\boldsymbol{\beta}) \sim N_p(\mathbf{b}, \mathbf{v})$, sendo p o número de parâmetros. A distribuição condicional completa de $\boldsymbol{\beta}$ é dada por

$$\begin{aligned} \boldsymbol{\beta}|\mathbf{z} &\sim N_p(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V}(\mathbf{v}^{-1}\mathbf{b} + \mathbf{X}^\top \mathbf{z}) \\ \mathbf{V} &= (\mathbf{v}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1} \end{aligned} \quad (3.4)$$

sendo $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)$. A distribuição condicional completa de cada elemento z_i é a normal truncada,

$$z_i|y_i, \boldsymbol{\beta} \propto \begin{cases} N(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)\mathbf{I}(z_i > 0) & \text{se } y_i = 1 \\ N(\mathbf{x}_i^\top \boldsymbol{\beta}, 1)\mathbf{I}(z_i \leq 0) & \text{caso contrário.} \end{cases} \quad (3.5)$$

Utilizaremos o transformador da inversa descrita por Devroye (1986) para gerar amostra da distribuição normal truncada apresentada em (3.5).

Com o uso da variável auxiliar, temos uma estrutura adequada para o uso de Amostrador de Gibbs. Para isso, gera-se sucessivas amostras das distribuições (3.4) e (3.5). Porém, em geral, observa-se alta correlação *a posteriori* entre $\boldsymbol{\beta}$ e \mathbf{z} , por isso utilizaremos a fatorização (3.6), sugestão feitas por Holmes e Held (2006). Com esta fatorização,

obtem-se a atualização conjunta dos coeficientes de regressão e das variáveis auxiliares.

$$\pi(\mathbf{z}, \boldsymbol{\beta} | \mathbf{y}) = \pi(\mathbf{z} | \mathbf{y}) \pi(\boldsymbol{\beta} | \mathbf{z}), \quad (3.6)$$

sendo $\pi(\boldsymbol{\beta} | \mathbf{z})$ dado em (3.4) e para \mathbf{z} condicional à \mathbf{y} vamos assumir que *a priori* de $\boldsymbol{\beta}$ segue uma distribuição normal n-variada, onde as médias podem assumir valores diferentes de zeros. Então temos que $\pi(\mathbf{z} | \mathbf{y})$ é dada por

$$\mathbf{z} | \mathbf{y} \propto N_n(\mathbf{X}\mathbf{b}, \mathbf{I}_n + \mathbf{X}\mathbf{v}\mathbf{X}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}),$$

sendo \mathbf{I}_n a matriz identidade de dimensão n e $\text{Ind}(\mathbf{y}, \mathbf{z})$ a função indicadora, dada por: $\text{Ind}(\mathbf{y}, \mathbf{z}) = \text{Ind}[(y_i = 1)I(z_i > 0) + (y_i = 0)I(z_i \leq 0)]$.

É complicado amostrar diretamente da distribuição normal n-variada truncada. Um forma eficiente de obter uma amostra desta distribuição é utilizar o algoritmo Gibbs. Neste caso, a distribuição condicional de cada elemento z_i , condicional a demais, é uma normal univariada truncada no zero dada por

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

sendo $m_i = \mathbf{x}_i^\top \mathbf{b} + (1 - h_{ii})^{-1} \sum_{k=1, k \neq i}^n h_{ik} (z_k - \mathbf{x}_k^\top \mathbf{b})$, $v_i = (1 - h_{ii})^{-1}$ e $h_{ik} = \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_k$, com \mathbf{V} definida em 3.4.

Um alternativa eficiente para calcular os parâmetros de localização m_i e v_i são dados por

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

em que z_i é o valor atual da i -ésima observação do vetor \mathbf{z} e \mathbf{B} dado em (3.4). Para cada atualização de algum z_i devemos recalculamos o \mathbf{B} através da seguinte relação

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

sendo \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 o i -ésimo vetor coluna da matriz $\mathbf{S} = \mathbf{V}\mathbf{X}^\top$.

O algoritmo desse procedimento encontra-se no Apêndice A.

3.2.2 Regressão Logística

No modelo de regressão logística pode-se assumir que o erro segue a distribuição logística padrão e assim obter o modelo de regressão logística. Porém, a distribuição *a posteriori* é desconhecida. Por isso, utilizaremos o resultado apresentado em Stefanski (1991), que mostra que a distribuição logística pode ser representada como uma escala de mistura de distribuição normal padrão em que a densidade de mistura está relacionada com a distribuição Kolmogorov-Smirnov (KS) (Apêndice C).

Considere a classe de distribuições para as funções simétricas definida por

$$\mathcal{F}_s = \left\{ F(\cdot) = \int_0^\infty \Phi\left(\frac{\cdot}{\delta}\right) dG(\delta), G \text{ é uma fda no intervalo } [0, \infty) \right\},$$

Considerando $F \in \mathcal{F}_s$, o modelo de regressão binária é dado por

$$p_i = \mathbb{P}(y_i = 1) = F(\mathbf{x}_i^\top \boldsymbol{\beta}) = \int_0^\infty \Phi\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\delta}\right) dG(\delta).$$

O modelo logito é dado quando $\delta = 4\psi^2$ e ψ tem a distribuição kolmogorov-Smirnov assintótica (KS). Cujas a função densidade é dado por

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

Utilizando a variável auxiliar $\mathbf{z} = (z_1, \dots, z_n)^\top$ podemos escrever o modelo da seguinte maneira

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

com $z_i | \boldsymbol{\beta}, \delta \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta(\psi))$, sendo $\delta(\psi) > 0$ para todo $\psi > 0$, $\delta(\cdot)$ é uma função bijetora e $g(\cdot)$ é uma densidade de mistura contínua.

Neste caso, a verossimilhança é dada por

$$L(\boldsymbol{\beta} | z, y) = \prod_{i=1}^n f(z_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \text{Ind} [(y_i = 1)I(z_i > 0) + (y_i = 0)I(z_i \leq 0)], \quad (3.7)$$

sendo $f(\cdot)$ a função de densidade de verossimilhança associada à $F(\cdot)$.

Adicionando mais uma variável $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$, o modelo logito pode ser representado como

$$\begin{aligned}
 y_i &= \begin{cases} 1 & \text{se } z_i > 0 \\ 0 & \text{se caso contrário} \end{cases} \\
 z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \\
 \epsilon_i &\sim N(0, \delta_i) \\
 \delta_i &= (2\psi_i)^2 \\
 \psi_i &\sim KS \\
 \boldsymbol{\beta} &\sim N_p(\mathbf{b}, \mathbf{v})
 \end{aligned} \tag{3.8}$$

sendo ψ_i , $i = 1, \dots, n$, são variáveis aleatórias que seguem a distribuição Kolmogorov-Smirnov (KS).

Considerando o modelo (3.8), a distribuição *a priori* de $\boldsymbol{\beta}$ é uma função de distribuição normal, $N_p(\mathbf{b}, \mathbf{v})$. A distribuição a posteriori de $\boldsymbol{\beta}$ condicionado em $\{\mathbf{z}, \boldsymbol{\delta}\}$ é dada por

$$\begin{aligned}
 \boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\delta} &\sim N_p(\mathbf{B}, \mathbf{V}) \\
 \mathbf{B} &= \mathbf{V}(\mathbf{v}^{-1}\mathbf{b} + \mathbf{X}^\top \mathbf{W}\mathbf{z}) \\
 \mathbf{V} &= (\mathbf{v}^{-1} + \mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \\
 \mathbf{W} &= \text{diag}(\delta_1^{-1}, \dots, \delta_n^{-1})
 \end{aligned} \tag{3.9}$$

sendo $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)$.

A função de distribuição condicional de cada elemento de z_i dado $\{y_i, \boldsymbol{\beta}, \delta_i\}$ é normal truncada com variância igual a δ_i ,

$$z_i | y_i, \boldsymbol{\beta}, \delta_i \propto \begin{cases} N(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta_i) I(z_i > 0) & \text{se } y_i = 1 \\ N(\mathbf{x}_i^\top \boldsymbol{\beta}, \delta_i) I(z_i \leq 0) & \text{caso contrário.} \end{cases} \tag{3.10}$$

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

Para construir o algoritmo de Gibbs utilizaremos as seguintes atualizações

1. gera-se $(\boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\delta})$, em seguida;
2. gera-se $(\mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{y})$, e por último;

3. gera-se $(\boldsymbol{\delta}|\mathbf{z}, \boldsymbol{\beta})$ e retorna para 1.

A construção de blocos no modelo logístico pode ser feito de duas maneiras. A primeira segue o mesmo procedimento que o modelo probito, atualizando $\{\mathbf{z}, \boldsymbol{\beta}\}$ conjuntamente. Ou seja,

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

Na equação (3.9), Holmes e Held (2006) assumem $\mathbf{b} = \mathbf{0}$. Generalizando para $\mathbf{b} \in \mathbb{R}^p$, a distribuição $\pi(\mathbf{z}|\boldsymbol{\delta}, \mathbf{y})$ é normal n-variada truncada dada por

$$\mathbf{z}|\mathbf{y}, \boldsymbol{\delta} \propto N_p(\mathbf{X}\mathbf{b}, \mathbf{W}^{-1} + \mathbf{X}\mathbf{v}\mathbf{X}^\top)\text{Ind}(\mathbf{y}, \mathbf{z}), \quad (3.11)$$

sendo $\text{Ind}(\mathbf{y}, \mathbf{z})$ função indicadora. Utiliza-se o algoritmo Gibbs para gerar amostra desta distribuição. As distribuições condicionais completas da distribuição (3.11) são dadas por

$$z_i|z_{-i}, \mathbf{y}, \boldsymbol{\delta} \propto \begin{cases} N(m_i, v_i)\text{I}(z_i > 0) & \text{se } y_i = 1 \\ N(m_i, v_i)\text{I}(z_i \leq 0) & \text{caso contrário.} \end{cases} \quad (3.12)$$

\mathbf{z}_{-i} denota o vetor de variáveis \mathbf{z} com a i -ésima variável removida. Os parâmetros m_i e v_i são calculados da seguinte maneira

$$m_i = \mathbf{x}_i^\top \mathbf{B} - \frac{h_i}{\delta_i - h_i}(z_i - \mathbf{x}_i^\top \mathbf{B}), \quad v_i = \frac{\delta_i^2}{\delta_i - h_i}, \quad (3.13)$$

sendo z_i e δ_i denotam os valores atuais de z_i e δ_i , respectivamente. \mathbf{B} é dado em (3.9) 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 (3.9). 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 de atualização anterior de \mathbf{B} e z_i , respectivamente, e \mathbf{s}_i denota o i -ésimo vetor coluna de matriz $\mathbf{S} = \mathbf{V}\mathbf{X}^\top$. O próximo passo deste algoritmo é amostrar $(\boldsymbol{\delta}|\boldsymbol{\beta}, \mathbf{z})$ usando o algoritmo dado no Apêndice A, algoritmo I, e recalculá-lo \mathbf{V} e \mathbf{B} a cada iteração.

A outra opção, denotado por Algoritmo II, é atualizar $\{\mathbf{z}, \boldsymbol{\delta}\}$ conjuntamente utilizando a fatorização 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^T s são logísticas truncadas independentes, dada por

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

sendo $\text{Lo}(a, b)$ denota a função densidade de distribuição logística com parâmetro de localização a e parâmetro de escala b .

Como a fda e a inversa da distribuição logística truncada tem uma forma analítica simples, isto facilita na implementação e eficiência do Algoritmo II. Porém, a correlação é menor no Algoritmo I. Entre os dois algoritmos de atualização conjunta apresentados para o modelo logístico, Holmes e Held (2006) recomendam o uso do Algoritmo I devido a simplicidade do código.

3.2.3 Modelo Probito-Assimétrico

Considere a classe de distribuições para funções de ligações assimétricas paramétricas definida por

$$\mathcal{F}_{\mathcal{A}} = \left\{ F_{\lambda}(z) = \int_0^{\infty} F(z - \lambda\omega) dG(\omega), \lambda \in \mathbb{R} \right\}, \quad (3.14)$$

sendo F uma função de distribuição simétrica em torno do zero com suporte nos reais e G a fda de uma distribuição assimétrica no intervalo $[0, \infty)$.

No modelo de regressão binária, quando o inverso da função de ligação $F_{\lambda} \in \mathcal{F}_{\mathcal{A}}$, é definido por

$$p_i = \mathbb{P}(y_i = 1) = F_{\lambda}(\mathbf{x}_i^T \boldsymbol{\beta}) = \int_0^{\infty} F(\mathbf{x}_i^T \boldsymbol{\beta} - \lambda\omega) dG(\omega),$$

sendo F e G definidos em (3.14).

O modelo probito-assimétrico definido pelo Azzalini (1985) é dado por $F_{\lambda} = \Phi_{SN}(u; \mu, \sigma^2, \lambda)$, sendo $\Phi_{SN}(\cdot; \mu, \sigma^2, \lambda)$, com os 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}_A definida em 3.14, então $F_\lambda(u) = \Phi(u; 0, 1 + \lambda^2, \lambda) = \Phi_{CDS}(u; \lambda)$, sendo $\Phi_{CDS}(\cdot; \lambda)$ a normal-assimétrica definida por Chen et al. (1999). Considerando que as distribuições *a priori* de $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e λ são independentes, o modelo probito-assimétrico pertencente à classe \mathcal{F}_A pode ser representado por

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0 \\ 0 & \text{se caso contrário} \end{cases} \\ z_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i^* \\ \epsilon_i^* &= -\lambda \omega_i + \epsilon_i \\ \epsilon_i &\sim N(0, 1) \\ \omega_i &\sim \text{HN}(0, 1) \\ \boldsymbol{\beta} &\sim N_p(\mathbf{b}, \mathbf{v}) \\ \lambda &\sim N(\alpha, \tau), \end{aligned} \tag{3.15}$$

sendo HN a distribuição normal positiva (Half normal).

Sob o modelo (3.15), a distribuição condicional completa de $\boldsymbol{\beta}$ é dada por

$$\begin{aligned} \boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\omega}, \lambda &\sim N_p(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V} [\mathbf{v}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z} + \lambda \boldsymbol{\omega})] \\ \mathbf{V} &= (\mathbf{v}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1}, \end{aligned} \tag{3.16}$$

sendo $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ a matriz de delineamento.

Na equação (3.17) temos a distribuição condicional completa de \mathbf{z} , com os componentes, z_i , para $i = 1, \dots, n$, variáveis aleatórias truncadas e independentes, ou seja:

$$z_i | y_i, \boldsymbol{\beta}, \lambda, \omega_i \propto \begin{cases} N(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda \omega_i, 1) I(z_i > 0) & \text{se } y_i = 1, \\ N(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda \omega_i, 1) I(z_i \leq 0) & \text{caso contrário.} \end{cases} \tag{3.17}$$

A distribuição condicional completa de $\boldsymbol{\omega}$ é normal n-variada truncada, sendo todos os componentes, ω_i , $i = 1, \dots, n$, variáveis aleatórias normais truncadas e independentes

$$\omega_i | z_i, \boldsymbol{\beta}, \lambda \propto N\left(-\frac{\lambda}{1 + \lambda^2} (z_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \frac{1}{1 + \lambda^2}\right) I(\omega_i > 0) \tag{3.18}$$

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

$$\begin{aligned}\lambda|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega} &\sim N(m, v) \\ m &= v [\tau^{-1}\alpha - \boldsymbol{\omega}^\top(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})] \\ v &= (\tau^{-1} + \boldsymbol{\omega}^\top\boldsymbol{\omega})^{-1}.\end{aligned}\tag{3.19}$$

3.3 Algoritmos de simulação

Existem várias maneiras de escolhas dos blocos para a implementação do Amostrador de Gibbs. Neste trabalho, vamos utilizar três diferentes maneiras de agrupamentos dos blocos, visando um algoritmo que forneça uma velocidade maior na convergência da cadeia. Deste modo, temos a atualização conjunta de $(\mathbf{z}, \boldsymbol{\beta})$, uma extensão dos modelos probito e logito. Outra opção é atualização conjunta de $(\mathbf{z}, \boldsymbol{\omega})$, similar ao Algoritmo II, Apêndice A, do modelo logístico. Além disso, podemos atualizar conjuntamente $(\mathbf{z}, \boldsymbol{\beta}, \lambda)$, uma representação similar ao de probito.

3.3.1 Atualização conjunta de $(\mathbf{z}, \boldsymbol{\beta})$

Utilizando a fatorização abaixo, obtemos a atualização de $\{\mathbf{z}, \boldsymbol{\beta}\}$ conjuntamente dado $\{\mathbf{y}, \boldsymbol{\omega}, \lambda\}$,

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

A distribuição $\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda)$ é dada em (3.16). No apêndice B mostramos que a distribuição de $\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda$ segue a seguinte densidade de probabilidade

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda) = C\phi_n(\mathbf{z}; \mathbf{X}\mathbf{b} - \lambda\boldsymbol{\omega}, I_n + \mathbf{X}\mathbf{v}\mathbf{X}^\top), \quad \mathbf{z} \in \mathfrak{R}(\mathbf{y}, \mathbf{z}),\tag{3.20}$$

sendo

$$C^{-1} = \bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \mathbf{X}\mathbf{b} - \lambda\boldsymbol{\omega}, I_n + \mathbf{X}\mathbf{V}\mathbf{X}^\top),$$

em que $\phi_n(\cdot, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a fdp de uma distribuição normal n -variada com vetor $\boldsymbol{\mu}$ e a matriz de covariâncias $\boldsymbol{\Sigma}$, e $\bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \boldsymbol{\mu}, \boldsymbol{\Sigma})$ é a função de distribuição acumulada na região $R(\mathbf{y}, \mathbf{z}) = \{\mathbf{z} = (z_1, \dots, z_n)^\top; z_i > 0 \text{ se } y_i = 1 \text{ ou } z_i \leq 0 \text{ se } y_i = 0\}$ de uma distribuição normal n -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$.

A seguir, apresentamos as distribuições condicionais completas (prova no Apêndice B) de $z_i | z_{-i}, y_i, \boldsymbol{\omega}$ onde utilizaremos um outro algoritmo de Gibbs.

$$z_i | \mathbf{z}_{-i}, y_i, \boldsymbol{\omega}, \lambda \propto \begin{cases} N(m_i, v_i) \mathbf{I}(z_i > 0) & \text{se } y_i = 0, \\ N(m_i, v_i) \mathbf{I}(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.21)$$

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

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

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 3.16. Alternativamente, podemos reescrever o parâmetro de localização m_i utilizando operações matriciais, ou seja

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

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 (3.16). Como \mathbf{B} é função das variáveis auxiliares z_i , devemos recalculá-lo para cada atualização de algum z_i utilizando a relação

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

sendo \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 de coluna de matriz $\mathbf{S} = \mathbf{V}\mathbf{X}^\top$.

Pode-se gerar amostra da distribuição *a posteriori* $\{\boldsymbol{\beta}, \lambda, \mathbf{z}, \boldsymbol{\omega}\}$ através das atualizações sucessivas

1. gera-se de $(\boldsymbol{\beta}, \mathbf{z} | \mathbf{y}, \boldsymbol{\omega}, \lambda)$ utilizando as distribuições (3.16) e (3.21), em seguida;
2. gera-se de $(\boldsymbol{\omega} | \mathbf{z}, \boldsymbol{\beta}, \lambda)$ utilizando (3.18), e por fim;
3. gera-se de $(\lambda | \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega})$ utilizando (3.19) e retorna a 1.

Denotamos por $HH\{\mathbf{z}, \boldsymbol{\omega}\}$ e o pseudo-código encontra-se no Apêndice A.

3.3.2 Atualização conjunta de $(\mathbf{z}, \boldsymbol{\omega})$

Utilizando a fatorização podemos escrever a distribuição *a posteriori* de $\{\mathbf{z}, \boldsymbol{\omega}\}$ dado $\{\boldsymbol{\beta}, \lambda\}$ da seguinte maneira

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

onde $\pi(\boldsymbol{\omega} | \mathbf{z}, \boldsymbol{\beta}, \lambda)$ é dada em (3.19). A distribuição de $\pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \lambda)$ é normal-assimétrica n-variada truncada, sendo que cada componente z_i , $i = 1, \dots, n$ tem distribuição normal-assimétrica truncada independentemente.

Ou seja,

$$z_i | y_i, \boldsymbol{\beta}, \lambda \propto \begin{cases} SN(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda) I(z_i > 0) & \text{se } y_i = 0, \\ SN(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda) I(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.22)$$

sendo $SN(\mu, \sigma^2, \lambda)$ a densidade da distribuição normal-assimétrica com parâmetro de localização μ , de escala σ^2 e de forma λ .

Denota-se por $HH(\mathbf{z}, \boldsymbol{\omega})$ e o pseudo-código é apresentado por Apêndice A.

3.3.3 Atualização conjunta de $(\mathbf{z}, \boldsymbol{\beta}, \lambda)$

Considere λ um coeficiente regressor associado as variáveis auxiliares $\boldsymbol{\omega}$. Fazendo $\mathbf{a}_i^\top = [\mathbf{x}_i^\top, \omega_i]$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda)^\top$, podemos representar esse modelo da seguinte maneira

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } z_i > 0, \\ 0 & \text{se caso contrário.} \end{cases} \\ z_i &= \mathbf{a}_i^\top \boldsymbol{\theta} + \epsilon_i \\ \epsilon_i &\sim N(0, 1) \\ \boldsymbol{\theta} &\sim N_{p+1}(\mathbf{b}, \mathbf{v}). \end{aligned} \quad (3.23)$$

Utilizando a fatorização abaixo, a atualização de $\{\mathbf{z}, \boldsymbol{\delta}\}$ é dada por

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

A distribuição de $\boldsymbol{\theta}$ condicional à $\{\mathbf{z}, \boldsymbol{\omega}\}$ é dada por

$$\begin{aligned}\boldsymbol{\theta}|\mathbf{z}, \boldsymbol{\omega} &\sim N_{p+1}(\mathbf{B}, \mathbf{V}) \\ \mathbf{B} &= \mathbf{V}(\mathbf{v}^{-1}\mathbf{b} + \mathbf{A}^\top \mathbf{z}) \\ \mathbf{V} &= (\mathbf{v}^{-1} + \mathbf{A}^\top \mathbf{A})^{-1}\end{aligned}\quad (3.24)$$

com $\mathbf{A} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top, \dots, \mathbf{a}_n^\top)$. A distribuição $\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega})$ é uma normal multivariada truncada (prova no Apêndice C) dada por

$$\mathbf{z}|\mathbf{y}, \boldsymbol{\omega} \propto N_n(\mathbf{X}\mathbf{b}, \mathbf{I}_n + \mathbf{A}\mathbf{v}\mathbf{A}^\top)\text{Ind}(\mathbf{y}, \mathbf{z}). \quad (3.25)$$

Para amostrar da distribuição (3.25), utilizaremos um novo algoritmo de Gibbs, através das seguintes distribuições condicionais (prova no Apêndice B)

$$z_i|z_{-i}, \mathbf{y}, \boldsymbol{\omega}, \lambda \propto \begin{cases} N(m_i, v_i)\text{I}(z_i > 0) & \text{se } y_i = 1, \\ N(m_i, v_i)\text{I}(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (3.26)$$

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

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

sendo \mathbf{B} como definida em (3.24), z_i denotando o valor atual de z_i , h_i é o i -ésimo elemento da diagonal da matriz $\mathbf{H} = \mathbf{A}\mathbf{V}\mathbf{A}^\top$ e \mathbf{V} definida em (3.24).

A atualização de \mathbf{B} é realizada a cada atualização de algum z_i utilizando a seguinte 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 da matriz $\mathbf{S} = \mathbf{V}\mathbf{A}^\top$.

Como a matriz $\mathbf{A} = [\mathbf{X}, \boldsymbol{\omega}]$ é função das variáveis auxiliares $\boldsymbol{\omega}$, a atualização de $\boldsymbol{\omega}$ é feita através da distribuição $\pi(\boldsymbol{\omega}|\mathbf{z}, \boldsymbol{\beta}, \lambda)$ apresentado em (3.18).

Denotamos por $HH(\mathbf{z}, \boldsymbol{\beta}, \lambda)$ e Algoritmo encontra-se no Apêndice A.

3.4 Identificação de Pontos Influentes

Considere o conjunto de dados $D = \{y_1, \dots, y_n\}$ e o conjunto de dados após ter sido retirado a i -ésima observação, $D_{(-i)} = \{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n\}$. Vamos denotar as densidades *a posteriori* para o vetor de parâmetros β , obtidas com cada um destes conjuntos de dados, por $\pi(\beta|D_{(-i)})$, respectivamente. A influência da observação y_i pode ser avaliada calculando a distância de Kullback-Leibler entre estas duas densidades *a posteriori*.

Esta medida, denotada por $KL(\pi, \pi_{(-i)})$, é a divergência de Kullback-Leibler entre π e $\pi_{(-i)}$ onde π denota a distribuição *a posteriori* para os dados completos e $\pi_{(-i)}$ é a distribuição *a posteriori* sem a i -ésima observação e é calculada por

$$KL(\pi, \pi_{(-i)}) = \int \pi(\beta|D) \log \left(\frac{\pi(\beta|D)}{\pi(\beta|D_{(-i)})} \right) d\beta, \quad (3.27)$$

sendo

$$\pi(\beta|D) = \frac{L(\beta|D)\pi(\beta)}{\int L(\beta|D)\pi(\beta)d\beta}$$

e

$$\pi(\beta|D_{(-i)}) = \frac{L(\beta|D_{(-i)})\pi(\beta)}{\int L(\beta|D_{(-i)})\pi(\beta)d\beta}$$

tal que

$$\begin{aligned} \frac{\pi(\beta|D)}{\pi(\beta|D_{(-i)})} &= \frac{L(\beta|D)}{L(\beta|D_{(-i)})} \frac{\int L(\beta|D_{(-i)})\pi(\beta)d\beta}{\int L(\beta|D)\pi(\beta)d\beta} \\ &= f(y_i|\beta) \frac{\int L(\beta|D_{(-i)})\pi(\beta)d\beta}{\int f(y_i|\beta)L(\beta|D_{(-i)})\pi(\beta)d\beta} \\ &= f(y_i|\beta) \frac{\int L(\beta|D_{(-i)})\pi(\beta)d\beta}{\int f(y_i|\beta)(\pi(\beta|D_{(-i)}) \int L(\beta|D_{(-i)})\pi(\beta)d\beta)d\beta} \\ &= f(y_i|\beta) \frac{\int L(\beta|D_{(-i)})\pi(\beta)d\beta}{\int L(\beta|D_{(-i)})\pi(\beta)d\beta \int f(y_i|\beta)\pi(\beta|D_{(-i)})d\beta} \\ &= f(y_i|\beta) \frac{1}{\int f(y_i|\beta)\pi(\beta|D_{(-i)})d\beta} \end{aligned} \quad (3.28)$$

Aplicando a função log na equação (3.28) obtemos

$$\log \left(\frac{\pi(\beta|D)}{\pi(\beta|D_{(-i)})} \right) = \log f(y_i|\beta) - \log \int f(y_i|\beta)\pi(\beta|D_{(-i)})d\beta \quad (3.29)$$

Denotando a densidade preditiva ordenada da observação y_i por

$$CPO_i = \int f(y_i|\boldsymbol{\beta})\pi(\boldsymbol{\beta}|D_{(-i)})d\boldsymbol{\beta}, \quad (3.30)$$

podemos reescrevê-la da forma

$$\begin{aligned} CPO_i &= \int f(y_i|\boldsymbol{\beta}) \frac{L(\boldsymbol{\beta}|D_{(-i)})\pi(\boldsymbol{\beta})}{\int L(\boldsymbol{\beta}|D_{(-i)})\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} d\boldsymbol{\beta} \\ &= \int \frac{L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})}{\int \frac{f(y_i|\boldsymbol{\beta})}{f(y_i|\boldsymbol{\beta})} L(\boldsymbol{\beta}|D_{(-i)})\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} d\boldsymbol{\beta} \\ &= \int \frac{L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})}{\int \frac{1}{f(y_i|\boldsymbol{\beta})} L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} d\boldsymbol{\beta} \\ &= \frac{1}{\int \frac{1}{f(y_i|\boldsymbol{\beta})} L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} \int L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})d\boldsymbol{\beta} \\ &= \frac{1}{\int \frac{1}{f(y_i|\boldsymbol{\beta})} L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} \frac{1}{\int \frac{1}{L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})} d\boldsymbol{\beta}} \\ &= \frac{1}{\int \frac{1}{f(y_i|\boldsymbol{\beta})} \frac{L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})}{\int L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}} d\boldsymbol{\beta}} \\ &= \frac{1}{\int \frac{1}{f(y_i|\boldsymbol{\beta})} \pi(\boldsymbol{\beta}|D)d\boldsymbol{\beta}}. \end{aligned} \quad (3.31)$$

Substituindo (3.29) e (3.30) em (3.27), tem-se

$$\begin{aligned} KL(\pi, \pi_{(-i)}) &= \int \log(f(y_i|\boldsymbol{\beta}))\pi(\boldsymbol{\beta}|D)d\boldsymbol{\beta} - \int \log(CPO_i)\pi(\boldsymbol{\beta}|D)d\boldsymbol{\beta} \\ &= E_{\pi(\boldsymbol{\beta}|D)} \{ \log f(y_i|\boldsymbol{\beta}) \} - \log(CPO_i). \end{aligned} \quad (3.32)$$

Considerando uma amostra gerada da densidade a posteriori $\pi(\boldsymbol{\beta}|D)$, dada por $\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(Q)}$, podemos estimar o efeito da observação y_i por

$$\widehat{KL}(\pi, \pi_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log f(y_i|\boldsymbol{\beta}^{(q)}),$$

em que a \widehat{CPO}_i é dada por

$$\widehat{CPO}_i = \left[\frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\boldsymbol{\beta}^{(q)})} \right]^{-1}.$$

Segundo Peng e Dey (1995) e Cho et al. (2009) a calibração de $KL(\pi, \pi_{(-i)})$ pode ser realizado resolvendo a equação KL para encontrar p_i ,

$$KL(\pi, \pi_{(-i)}) = KL(B(0.5), B(p_i)) = \frac{-\log \{4p_i(1-p_i)\}}{2} \quad (3.33)$$

em que, $B(p)$ é uma função de distribuição de Bernoulli com probabilidade de sucesso p . Resolvendo a equação (3.33) temos que

$$p_i = \frac{1}{2} \left\{ 1 + \sqrt{1 - \exp[-2KL(\pi, \pi_{(-i)})]} \right\} \quad (3.34)$$

Esta equação implica que $0.5 \leq p_i \leq 1$. Então se $p_i \gg 0.5$ então a i -ésima observação é considerado ponto influente.

3.5 Critérios Para Comparação de Modelos

O critério proposto Akaike (1974) baseia-se na verossimilhança penalizada pelo número de parâmetros do modelo, $AIC = -2 \sum_{i=1}^n \log L(\hat{\beta}; y_i) + 2p$, enquanto que o critério de informação bayesiano (BIC) proposto por Schwarz et al. (1978) pondera o tamanho amostral com $BIC = -2 \sum_{i=1}^n \log L(\hat{\beta}; y_i) + p \log(n)$. Os critérios de seleção, no contexto bayesiano, são obtidos através de uma extensão considerando a densidade *a posteriori* dos parâmetros do modelo.

$$EAIC = -2E \left(\sum_{i=1}^n \log L(\hat{\beta}; y_i) \right) + 2p, \quad (3.35)$$

$$EBIC = -2E \left(\sum_{i=1}^n \log L(\hat{\beta}; y_i) \right) + p \log(n), \quad (3.36)$$

sendo p é o número de parâmetros do modelo, n é o tamanho da amostral. Ambos os critérios, EAIC e EBIC, indicam os melhores modelos quanto menor for o valor obtido.

3.6 Medidas de Eficiência

Através de uma medida padrão para descrição do erro, é possível verificar se o método estudado é adequado e, ainda, estabelecer uma escala de comparação entres diferentes modelos ou metodologias.

Considere uma cadeia de tamanho N , réplica de tamanho m e um vetor de parâmetros β . Podemos definir $\bar{\beta}_i$ da seguinte maneira

$$\bar{\beta}_i = \frac{1}{N} \sum_{j=1}^N \hat{\beta}_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, N, \quad (3.37)$$

sendo $\hat{\beta}_{ij}$ o j -ésimo vetor da i -ésima observação de uma amostra de tamanho m da distribuição *a posteriori* de β e $\bar{\beta}_i$ é a estimativa obtida na réplica no instante i .

Uma das medidas do erro mais simples é o erro médio (EM) e é definido pela equação (3.38). Este erro é simples e fácil de ser calculado, porém, não é uma medida confiável e tendem a ser pequeno pela possibilidade de ocorrência de erros negativos e positivos, que acabam anulando durante a somatória.

$$\text{EM} = \frac{1}{N} \sum_{i=1}^N (\beta - \bar{\beta}_i). \quad (3.38)$$

Para resolver o problema de cancelamento dos erros de sinais opostos, ocorrido no EM, podemos optar por outra medida, o erro quadrático médio (EQM) que eleva os erros ao quadrado. O EQM tem sempre valor positivo e quanto mais próximo de zero, maior a qualidade dos valores medidos ou estimados. A fórmula para o cálculo do EQM é apresentado pela equação (3.39).

$$\text{EQM} = \frac{1}{N} \sum_{i=1}^N (\beta - \bar{\beta}_i)^2. \quad (3.39)$$

Mesmo corrigindo o problema do sinal, o EQM é falho no que se refere a magnitude, ela depende diretamente da escala dos dados, e isto pode dificultar nas análises. Por este motivo, a utilização de medidas percentuais é a mais adequada na avaliação de medidas de eficiência. Uma das medidas dos erros percentuais, é o erro percentual absoluto médio MAPE (*mean absolute percentage error*), definido pela equação (3.40). Ela é dado pela média de todos os erros absolutos percentuais, fornecendo uma indicação do tamanho médio do erro, expresso como uma porcentagem do valor observado, independentemente do erro ser positivo ou negativo. Porém, apresenta problema quando o valor β for igual a 0, devido à sua utilização no denominador das equações, além disso, esta medida apresenta problema de simetria quando o valor estimado é muito maior ou menor que o verdadeiro, ou quando os valores das previsões são próximos de zero.

Como forma de contornar a assimetria existente no MAPE. Makridakis e Hibon (1995) propuseram a medição da precisão das previsões, através do cálculo do erro percentual absoluto médio simétrico (SMAPE ou sMAPE, *symmetric mean absolute percentage error*), apresentado na equação (3.41).

$$\text{MAPE} = \frac{100\%}{N} \sum_{j=1}^n \left| \frac{\beta - \widehat{\beta}_i}{\beta} \right|. \quad (3.40)$$

$$\text{SMAPE} = \frac{100\%}{N} \sum_{j=1}^n \frac{|\beta - \widehat{\beta}_i|}{(|\beta| + |\widehat{\beta}_i|)/2}. \quad (3.41)$$

3.7 Critério de Convergência

O teste de Geweke divide a cadeia de Markov em duas partes e testa se os valores da primeira parte da cadeia são iguais em média aos valores da segunda parte. Se a distribuição é estacionária, espera-se que a média da primeira parte seja igual a segunda. Em geral, a primeira parte corresponde a 10% das iterações após o período de aquecimento e a segunda parte são os últimos 50% das iterações da cadeia. Seja β^t a cadeia de Markov, dividindo em duas subsequências, obtemos $\beta_1^t : t = 1, \dots, n_1$ e $\beta_2^t : t = n_a, \dots, n$, sendo $1 < n_1 < n_a < n$ e $n_2 = n - n_a + 1$. Então as médias das subsequências são $\bar{\beta}_1 = \sum_{t=1}^{n_1} \frac{\beta^t}{n_1}$ e $\bar{\beta}_2 = \sum_{t=n_a}^n \frac{\beta^t}{n_2}$.

Com as suas respectivas variâncias estimados, S_1^2 e S_2^2 . A estatística do teste é dada por

$$Z_n = \frac{\bar{\beta}_1 - \bar{\beta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Admitindo-se que $\frac{n_1}{N}$ e $\frac{n_2}{N}$ são fixos. Quando $N \rightarrow \infty$ a distribuição da estatística converge para uma $N(0, 1)$. Se o valor p obtido for maior que o nível de significância pré-fixado, então temos evidência para concluir que a distribuição *a posteriori* convergiu.

Podemos concluir, a abordagem Bayesiana aparentemente exige um maior esforço computacional devido aos algoritmos de MCMC, porém é uma abordagem bastante flexível para resolver o problema de inferência com diferentes modelos, assim como não

é necessário de grandes amostras como é o caso da teoria clássica com as propriedades assintóticas. Nos estudos de simulação veremos a viabilidade destas abordagens.

3.8 Considerações Finais

Podemos concluir, a abordagem bayesiana aparentemente exige um maior esforço computacional devido aos algoritmos de MCMC, porém é uma abordagem bastante flexível para resolver o problema de inferência com diferentes modelos, assim como não é necessário de grandes amostras como é o caso da teoria clássica com as propriedades assintóticas. Nos estudos de simulação veremos a viabilidade destas abordagens.

Resultados de Simulações

Na inferência clássica e bayesiana sem o uso de variáveis auxiliares, os dados (y_i, \mathbf{x}_i) , em que y_i é o i -ésimo elemento do vetor da variável resposta e \mathbf{x}_i é um vetor coluna formado pelo i -ésima observação das p variáveis explicativas, foram gerados a partir de 4.1, em que, p_i é calculado de acordo com a função de ligação. As funções de ligações consideradas neste estudos formam probito, logito, complemento log-log, transformação box-cox e probito assimétrico.

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{se } u_i \leq p_i, \\ 0, & \text{caso contrário.} \end{cases} \\ u_i &\sim \text{Uniforme}(0, 1) \\ p_i &\sim g^{-1}(\eta_i) \\ \mathbf{x}_i &\sim \text{Uniforme}(0, 1). \end{aligned} \tag{4.1}$$

Para o caso de inferência bayesiana com o uso de variáveis auxiliares, trabalhamos com um conjunto de dados que foram simulados a partir das equações (3.3), (3.8) e (3.15) para as funções de ligações probito, logito e probito-assimétrico, respectivamente. A matriz de covariável, \mathbf{X} , foi gerada a partir da distribuição uniforme com média zero e variância um.

Os algoritmos foram implementados no programa estatístico R Team et al. (2013) por ser um software gratuito e por ter vários pacotes implementados úteis para este trabalho, como o pacote `sn` de Azzalini (2011) para gerar amostras de uma variável normal-assimétrica e para os cálculos das distribuições acumulada e acumulada inversa da distribuição normal-assimétrica, os pacotes `coda` Plummer et al. (2006) e `stats` para verificar a convergência e a autocorrelação, respectivamente, das cadeias.

4.1 Inferência Clássica

Para o estudo de percentual de cobertura, o processo iterativo para obter $\hat{\boldsymbol{\beta}}$ pode ser expresso como um processo iterativo de mínimos quadrados ponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{K}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

em que, \mathbf{K} é a matriz de informação de Fisher dada por: $\mathbf{K} = \mathbf{X}^\top \mathbf{W} \mathbf{X}$ e $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})$.

Assintoticamente, $n \rightarrow \infty$, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N_p(\mathbf{0}, (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1})$. O intervalo de confiança(IC) de 95% para $\boldsymbol{\beta}$ é dado por

$$IC = \hat{\boldsymbol{\beta}} \pm 1.96 \sqrt{Var(\hat{\boldsymbol{\beta}})}, \quad (4.2)$$

em que, $Var(\hat{\boldsymbol{\beta}}) = diag[(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}]$.

Temos interesse em saber se o percentual de cobertura do intervalo de confiança assintótico está de acordo com o esperado na teoria. Para isso, consideramos amostras de tamanhos 500, 1000 e 1500 e réplica de tamanho 1000.

Geramos os dados a partir da equação (4.1) e ajustamos o modelo. Depois calculamos o intervalo de confiança assintótico utilizando a equação (4.2).

Nas Tabelas 4.1 mostramos o resultado de simulação do percentual de cobertura para funções de ligações probito, logito, complemento log-log, transformação box-cox e probito-assimétrico. Como podemos verificar, aproximadamente 95% dos intervalos construídos contém o verdadeiro valor do parâmetro.

Tabela 4.1: Percentual de cobertura utilizando o método de máxima verossimilhança.

Função	θ	n		
		500	1000	1500
Logito	$\beta_0 = -1.5$	96.3	95.7	95.6
	$\beta_1 = 2.5$	96.2	95.2	95.8
	$\beta_2 = 4.5$	96.1	95.6	95.3
Probito	$\beta_0 = -2.0$	94.8	95.8	95.0
	$\beta_1 = 3.5$	95.7	97.0	95.1
	$\beta_2 = 4.0$	95.0	95.4	95.0
Cloglog	$\beta_0 = -3.5$	96.4	95.5	96.9
	$\beta_1 = 3.0$	95.6	96.2	95.0
	$\beta_2 = 2.5$	95.8	95.4	96.2
Transf. box-cox	$\lambda = 0.5$	99.9	99.9	99.9
	$\beta_0 = -1.5$	97.7	98.2	98.3
	$\beta_1 = 2.5$	97.7	98.2	98.3
	$\beta_2 = 1.2$	95.5	95.9	95.2
Probito-Assimétrico	$\lambda = 1.5$	99.9	99.9	99.9
	$\beta_0 = -1.5$	95.0	95.1	95.4
	$\beta_1 = 2.5$	95.1	95.2	95.5
	$\beta_2 = 1.5$	95.2	94.9	95.1

Os critérios AIC e BIC, geralmente, são utilizados para selecionar modelos e variáveis explicativas. Neste estudo de simulação queremos mostrar que estes critérios também podem ser utilizados para selecionar função de ligação. Deste modo, consideramos os valores de parâmetros, $\beta = (-1.5, 2.5, 4.5)$ para a função de ligação logito, $\beta = (-2.0, 3.5, 4.0)$ para probito e $\beta = (-3.5, 3.0, 2.5)$ para a função complemento log-log. Os dados foram gerados a partir da equação (4.1) e ajustado para diferentes funções de ligações e contabilizamos as funções que apresentaram menor valor de AIC e BIC para as mil réplicas.

Podemos ver na Tabela 4.2 que o valor percentual de escolha é maior quando o valor da variável resposta é gerado pela mesma função de ligação que foi ajustado.

Tabela 4.2: Percentual de escolha utilizando os critérios AIC e BIC.

n	Função	AIC			BIC		
		Logito	Probita	Cloglog	Logito	Probita	Cloglog
500	Logito	60.1	25.1	14.8	65.5	28.7	5.8
	Probita	17.8	73.8	8.4	23.1	74.1	2.8
	Cloglog	6.9	10.8	82.3	0.6	11.7	87.7
1000	Logito	70.3	26.7	3.0	70.1	27.0	2.9
	Probita	17.2	79.2	3.4	22.4	75.4	2.2
	Cloglog	3.3	3.3	93.4	0.1	9.3	90.6
1500	Logito	76.1	23.4	0.5	70.4	28.4	1.2
	Probita	14.2	84.5	1.3	21.4	76.4	2.2
	Cloglog	1.9	1.7	96.4	0.0	7.7	92.3

4.2 Inferência bayesiana

No método bayesiano, trabalhamos com amostras de tamanhos 50, 100, 150 e 250, consideramos 50 mil iterações sendo 10 mil período de aquecimento, um salto de 30 em 30 e réplica de 500.

Para cada uma das cadeias, monitoramos os gráficos da função de autocorrelação (FAC). Aplicamos o teste Geweke, Geweke et al. (1991), nas iterações após o período de aquecimento para verificar a convergência da cadeia.

Na inferência bayesiana sem o uso variáveis auxiliares, a distribuição *a priori* usada para o vetor de parâmetros $\beta = (\beta_0, \beta_1, \beta_2)^\top$ é $\beta \sim N_3(\mathbf{0}, 25\mathbf{I}_3)$.

Para as funções de ligações simétricos, probito e logito, apresentamos os resultados das simulações nas Tabelas 4.3 e 4.4 utilizando inferência bayesiana com e sem variáveis auxiliares. Para as funções assimétricas, complemento log-log, transformação box-cox e probito-assimétrico, utilizando inferência bayesiana sem o uso de variáveis auxiliares, os resultados encontram-se na Tabela 4.5. Por fim, para a função de ligação probito-assimétrico e utilizando inferência bayesiana com o uso de variáveis auxiliares, apresentamos os resultados na Tabela 4.6.

Observa-se que na inferência bayesiana sem o uso de variáveis auxiliares, para amostras pequenas obtemos resultados melhores para as funções assimétricas, transformação

box-cox e probito-assimétrico. No caso de inferência bayesiana com o uso de variáveis auxiliares, as funções logito e probito-assimétrico apresentam melhores resultados.

Para comparar os resultados das duas metodologias bayesianas, com e sem o uso de variáveis auxiliares, apresentamos os resultados obtidos com os cálculos de EQM, MAPE e SMAPE para as funções probito, logito, complemento log-log, transformação box-cox e probito-assimétrico, nas Tabelas 4.7, 4.8, 4.9 e 4.10, respectivamente. Para a função logito apresentamos os algoritmos I e II, e para a função probito-assimétrico apresentamos os algoritmos I, II e III. Os algoritmos I e III das funções logito e probito-assimétrico, respectivamente, apresentam melhores resultados.

Tabela 4.3: Resultados das estimativas usando a função de ligação probito, utilizando os métodos com e sem o uso de variáveis auxiliares.

		Probito							
		Sem Variável Auxiliar				Com Variável Auxiliar			
n	Parâmetros	Média	Mediana	Moda	DP ¹	Média	Mediana	Moda	DP ¹
50	$\beta_0 = -1.5$	-1.329	-1.323	-1.296	0.318	-1.391	-1.301	-1.218	0.421
	$\beta_1 = 2.0$	1.724	1.747	1.875	0.427	1.821	1.713	1.664	0.587
	$\beta_2 = 1.5$	1.657	1.657	1.767	0.460	1.606	1.793	1.662	0.441
100	$\beta_0 = -1.5$	-1.400	-1.399	-1.325	0.286	-1.400	-1.399	-1.325	0.286
	$\beta_1 = 2.0$	1.865	1.854	1.772	0.424	1.865	1.854	1.772	0.424
	$\beta_2 = 1.5$	1.622	1.605	1.564	0.369	1.622	1.605	1.564	0.369
150	$\beta_0 = -1.5$	-1.454	-1.439	-1.404	0.242	-1.454	-1.439	-1.404	0.242
	$\beta_1 = 2.0$	1.919	1.924	1.903	0.314	1.919	1.924	1.903	0.314
	$\beta_2 = 1.5$	1.542	1.542	1.631	0.375	1.542	1.542	1.631	0.375
250	$\beta_0 = -1.5$	-1.490	-1.473	-1.404	0.236	-1.490	-1.473	-1.404	0.236
	$\beta_1 = 2.0$	1.932	1.937	1.899	0.280	1.932	1.937	1.899	0.280
	$\beta_2 = 1.5$	1.513	1.483	1.554	0.320	1.513	1.483	1.554	0.320

DP¹: Desvio Padrão

Tabela 4.4: Resultados das estimativas utilizando a função de ligação logito, usando os métodos sem e com o uso variáveis auxiliares.

		Logito											
		Sem Variáveis Auxiliares				Com Variáveis Auxiliares							
						Algoritmo I				Algoritmo II			
n	β	Média	Mediana	Moda	DP ¹	Média	Mediana	Moda	DP ¹	Média	Mediana	Moda	DP ¹
50	$\beta_0 = -1.5$	-1.625	-1.592	-1.481	0.823	-1.493	-1.489	-1.502	0.103	-1.517	-1.509	-1.512	0.330
	$\beta_1 = 2.0$	2.230	2.118	2.017	1.216	1.995	1.994	2.002	0.131	1.986	1.983	1.982	0.198
	$\beta_2 = 1.5$	1.614	1.525	1.398	1.082	1.496	1.495	1.493	0.040	1.496	1.498	1.496	0.078
100	$\beta_0 = -1.5$	-1.603	-1.598	-1.463	0.606	-1.505	-1.515	-1.552	0.228	-1.505	-1.507	-1.506	0.219
	$\beta_1 = 2.0$	2.137	2.108	2.011	0.739	2.003	2.008	2.017	0.063	1.986	1.983	1.982	0.198
	$\beta_2 = 1.5$	1.610	1.602	1.381	0.809	1.501	1.501	1.498	0.029	1.496	1.498	1.497	0.055
150	$\beta_0 = -1.5$	-1.405	-1.360	-1.425	0.402	-1.503	-1.492	-1.487	0.147	-1.504	-1.488	-1.668	0.442
	$\beta_1 = 2.0$	1.930	1.885	1.819	0.528	1.997	2.000	2.006	0.095	2.005	2.002	2.001	0.134
	$\beta_2 = 1.5$	1.472	1.464	1.523	0.45	1.500	1.500	1.498	0.032	1.507	1.510	1.506	0.055
250	$\beta_0 = -1.5$	-1.553	-1.529	-1.425	0.373	-1.499	-1.501	-1.534	0.190	-1.493	-1.475	-1.491	0.219
	$\beta_1 = 2.0$	2.087	2.067	1.819	0.531	2.001	1.999	2.011	0.113	1.993	1.990	1.991	0.134
	$\beta_2 = 1.5$	1.530	1.517	1.523	0.483	1.499	1.500	1.502	0.055	1.498	1.506	1.579	0.136

DP¹: Desvio Padrão

Tabela 4.5: Resultados das estimativas utilizando as funções de ligações assimétricas, complemento log-log, transformação box-cox e probito-assimétrico, usando o método sem o uso variáveis auxiliares.

Funções		Complemento log-log				Transformação box-box				Probit-Assimétrico			
n	θ	Média	Mediana	Moda	DP	Média	Mediana	Moda	DP	Média	Mediana	Moda	DP
50	$\beta_0 = -1.5$	-1.848	-1.742	-1.538	0.746	-1.516	-1.519	-1.580	0.182	-1.513	-1.512	-1.476	0.127
	$\beta_1 = 2.0$	2.472	2.269	2.011	1.092	1.990	1.989	1.978	0.172	2.004	2.004	2.019	0.132
	$\beta_2 = 1.5$	1.821	1.760	1.682	0.878	1.478	1.488	1.492	0.179	1.514	1.519	1.507	0.131
	λ					0.599	0.600	0.601	0.018	0.999	1.000	1.000	0.011
100	$\beta_0 = -1.5$	-1.617	-1.589	-1.428	0.472	-1.511	-1.518	-1.493	0.170	-1.506	-1.514	-1.523	0.103
	$\beta_1 = 2.0$	2.195	2.166	2.009	0.617	2.008	2.020	2.019	0.171	1.997	1.994	2.007	0.099
	$\beta_2 = 1.5$	1.580	1.544	1.666	0.435	1.512	1.510	1.527	0.159	1.505	1.507	1.505	0.103
	λ					0.598	0.598	0.597	0.011	1.000	1.001	0.999	0.009
150	$\beta_0 = -1.5$	-1.593	-1.600	-1.618	0.367	-1.489	-1.489	-1.478	0.168	-1.497	-1.495	-1.460	0.097
	$\beta_1 = 2.0$	2.107	2.108	2.067	0.473	1.995	2.004	1.981	0.164	2.002	2.005	1.986	0.099
	$\beta_2 = 1.5$	1.563	1.562	1.521	0.353	1.507	1.517	1.545	0.158	1.503	1.500	1.488	0.097
	λ					0.597	0.599	0.600	0.006	0.999	0.999	0.998	0.008
250	$\beta_0 = -1.5$	-1.570	-1.550	-1.390	0.274	-1.500	-1.501	-1.540	0.162	-1.502	-1.499	1.501	0.091
	$\beta_1 = 2.0$	2.077	2.053	1.905	0.365	2.004	2.014	2.008	0.161	1.998	2.003	1.995	0.095
	$\beta_2 = 1.5$	1.559	1.517	1.521	0.442	1.509	1.515	1.524	0.151	1.501	1.502	1.508	0.096
	λ					0.590	0.592	0.601	0.006	1.000	1.000	1.000	0.008

Os valores de λ são: 0.5 e 1.0 para as funções transformação box-cox e probito-assimétrica, respectivamente.

Tabela 4.6: Resultados obtidos com a estimação da função de ligação probito-assimétrico, utilizando o método com o uso de variáveis auxiliares.

Probito-assimétrico													
n	Parâmetros	Algoritmo (z, β)				Algoritmo (β, w)				Algoritmo (z, β, λ)			
		Média	Mediana	Moda	DP	Média	Mediana	Moda	DP	Média	Mediana	Moda	DP
50	$\beta_0 = -1.5$	-1.482	-1.485	-1.487	0.742	-1.362	-1.343	-1.268	-1.417	-1.572	-1.597	-1.523	0.739
	$\beta_1 = 2.0$	2.012	2.001	2.020	0.711	2.053	2.007	1.996	1.734	1.946	1.942	2.068	0.711
	$\beta_2 = 1.5$	1.517	1.519	1.455	0.725	1.576	1.526	1.523	1.694	1.517	1.524	1.467	0.713
	$\lambda = 1.0$	0.003	0.002	0.008	0.061	1.694	1.696	1.696	1.691	1.014	0.999	0.993	0.740
100	$\beta_0 = -1.5$	-1.531	-1.517	-1.428	0.729	-1.362	-1.435	-1.463	1.033	-1.503	-1.501	-1.640	0.716
	$\beta_1 = 2.0$	1.992	1.995	2.040	0.695	2.053	2.045	2.093	1.000	1.983	1.963	1.961	0.733
	$\beta_2 = 1.5$	1.483	1.503	1.438	0.722	1.576	1.587	1.396	0.983	1.504	1.506	1.644	0.705
	$\lambda = 1.0$	0.007	0.007	0.010	0.083	1.694	1.692	1.692	0.103	0.995	1.011	1.154	0.737
150	$\beta_0 = -1.5$	-1.519	-1.502	-1.409	0.713	-1.431	-1.438	-1.479	0.992	-1.503	-1.514	-1.478	0.713
	$\beta_1 = 2.0$	1.994	1.984	1.945	0.714	2.051	2.054	2.102	0.978	2.013	2.009	2.054	0.710
	$\beta_2 = 1.5$	1.492	1.467	1.432	0.704	1.558	1.550	1.569	0.973	1.499	1.490	1.544	0.719
	$\lambda = 1.0$	0.009	0.009	0.016	0.102	1.686	1.687	1.686	0.072	0.995	1.001	1.039	0.710
250	$\beta_0 = -1.5$	-1.483	-1.475	-1.502	0.727	-1.451	-1.447	-1.494	0.964	-1.474	-1.476	-1.487	0.705
	$\beta_1 = 2.0$	2.000	2.032	2.011	0.718	2.026	2.016	2.022	0.994	2.000	1.994	2.004	0.706
	$\beta_2 = 1.5$	1.501	1.511	1.429	0.720	1.536	1.530	1.489	0.972	1.498	1.499	1.504	0.613
	$\lambda = 1.0$	0.020	0.018	0.020	0.147	1.676	1.677	1.682	0.044	0.999	0.964	0.927	0.734

Tabela 4.7: Critério de precisão para a função probito, utilizando o método com e sem o uso de variáveis auxiliares.

		Probit							
Parâmetros	Critérios de Precisão	Sem Variável Auxiliar				Com Variável Auxiliar			
		n				n			
		50	100	150	250	50	100	150	250
β_0	EQM	0.130	0.092	0.058	0.057	0.130	0.003	0.003	0.001
	MAPE	20.287	16.173	12.941	12.541	4.131	3.333	2.950	1.929
	SMAPE	21.944	17.479	13.120	12.964	4.144	3.329	2.944	2.027
β_1	EQM	0.258	0.198	0.105	0.083	0.258	0.002	0.001	0.001
	MAPE	20.007	18.233	12.555	11.330	3.836	1.637	1.483	1.281
	SMAPE	23.490	19.685	13.216	11.806	3.869	1.635	1.482	1.282
β_2	EQM	0.236	0.142	0.151	0.102	0.236	0.002	0.001	0.001
	MAPE	25.893	20.715	20.030	16.807	3.836	2.157	1.973	1.762
	SMAPE	24.477	20.063	19.548	16.958	2.662	2.160	1.976	1.763

Tabela 4.8: Resultados obtidos com os critérios de precisões para a função logito, usando os métodos sem e com o uso de variáveis auxiliares.

		Logito											
		Sem Variáveis Auxiliares				Com Variáveis Auxiliares							
						Algoritmo I				Algoritmo II			
Parâmetros	Critérios	n				n				n			
		50	100	150	250	50	100	150	250	50	100	150	250
β_0	EQM	0.692	0.377	0.214	0.141	0.036	0.021	0.010	0.010	0.193	0.109	0.048	0.022
	MAPE	42.375	32.628	24.814	19.614	10.209	7.868	2.815	1.147	11.815	7.999	5.557	3.221
	SMAPE	42.982	32.887	24.378	19.293	10.310	7.853	2.909	1.278	11.909	8.172	5.586	3.110
β_1	EQM	1.529	0.565	0.327	0.289	0.012	0.009	0.004	0.002	0.061	0.018	0.039	0.018
	MAPE	48.554	29.435	24.911	21.363	4.563	3.781	2.517	1.088	8.103	5.413	3.117	2.259
	SMAPE	47.919	29.276	23.268	20.940	4.563	3.792	2.517	1.112	8.204	5.433	3.034	2.391
β_2	EQM	1.182	0.666	0.344	0.234	0.003	0.001	0.0008	0.0008	0.018	0.006	0.003	0.002
	MAPE	57.228	43.318	35.224	26.108	2.936	2.179	1.537	1.197	4.191	2.936	2.111	1.449
	SMAPE	59.401	44.806	34.010	26.768	2.948	2.186	1.536	1.132	4.212	2.948	2.112	1.526

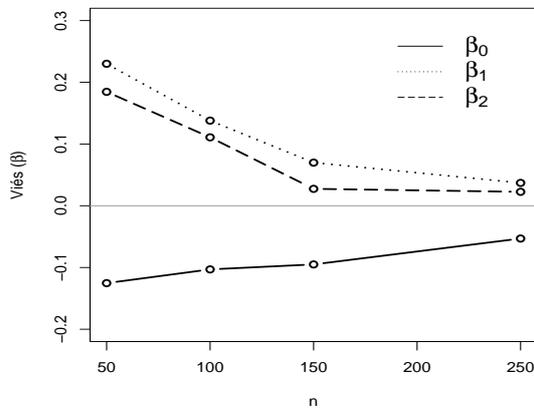
Tabela 4.9: Resultados obtidos usando os critérios de precisões para as funções: complemento log-log, transformação box-cox e Probit-assimétrico, utilizando o método sem o uso de variáveis auxiliares.

Parâmetros	Critérios	Complemento log-log				Transformação box-cox				Probit-assimétrico			
		n				n				n			
		50	100	150	250	50	100	150	250	50	100	150	250
β_0	EQM	0.677	0.236	0.143	0.080	0.033	0.029	0.028	0.026	0.902	0.010	0.009	0.008
	MAPE	41.832	24.950	20.355	14.868	9.847	9.081	8.975	8.524	6.863	5.459	5.234	4.844
	SMAPE	35.811	23.529	19.693	14.281	9.917	9.012	8.972	8.654	6.871	5.482	5.190	4.842
β_1	EQM	1.415	0.419	0.235	0.139	0.029	0.029	0.026	0.025	0.017	0.009	0.009	0.009
	MAPE	42.998	24.840	19.187	15.024	6.943	6.876	6.553	6.530	5.224	3.952	3.938	3.759
	SMAPE	35.984	23.092	18.564	14.562	6.953	6.908	6.603	6.548	5.242	3.944	3.942	3.765
β_2	EQM	0.873	0.331	0.195	0.128	0.032	0.025	0.024	0.023	0.017	0.010	0.009	0.009
	MAPE	47.646	29.630	23.450	18.719	9.676	8.584	8.441	8.088	6.902	5.472	5.157	5.073
	SMAPE	42.798	29.829	23.020	18.345	9.638	8.593	8.482	8.268	6.908	5.487	5.125	5.068
λ	EQM					0.070	0.009	0.009	0.008	0.902	0.078	0.648	0.006
	MAPE					30.725	19.763	19.598	18.093	5.760	5.157	4.641	3.672
	SMAPE					36.307	17.978	17.831	16.540	5.760	5.710	4.641	3.688

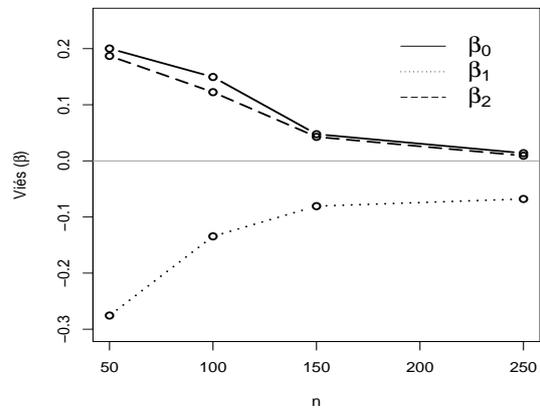
Tabela 4.10: Resultados obtidos utilizando os critérios de precisões para a função de ligação probito-assimétrico, usando método com o uso de variáveis auxiliares.

Probito-assimétrico										
n	Parâmetros	Algoritmo (z, β)			Algoritmo (β, w)			Algoritmo (z, β, λ)		
		EQM	MAPE	SMAPE	EQM	MAPE	SMAPE	EQM	MAPE	SMAPE
50	β_0	0.545	0.390	0.217	1.085	0.547	0.321	0.532	0.385	0.208
	β_1	0.538	0.288	0.155	1.000	0.399	0.217	0.515	0.286	0.153
	β_2	0.525	0.388	0.215	0.968	0.526	0.391	0.518	0.382	0.211
	λ	0.995	0.995	0.948	0.584	0.695	0.357	0.548	0.586	0.324
100	β_0	0.528	0.384	0.213	1.039	0.532	0.300	0.518	0.381	0.208
	β_1	0.509	0.286	0.154	0.991	0.398	0.217	0.508	0.285	0.152
	β_2	0.521	0.384	0.215	0.959	0.526	0.287	0.517	0.381	0.210
	λ	0.991	0.992	0.933	0.584	0.692	0.356	0.543	0.585	0.319
150	β_0	0.509	0.377	0.209	0.990	0.531	0.297	0.503	0.372	0.207
	β_1	0.506	0.285	0.151	0.960	0.390	0.212	0.504	0.282	0.150
	β_2	0.519	0.384	0.213	0.951	0.525	0.285	0.509	0.380	0.210
	λ	0.991	0.989	0.918	0.577	0.688	0.355	0.538	0.577	0.314
250	β_0	0.497	0.379	0.206	0.932	0.524	0.297	0.492	0.376	0.204
	β_1	0.499	0.283	0.153	0.920	0.387	0.210	0.482	0.281	0.149
	β_2	0.495	0.380	0.209	0.950	0.513	0.273	0.497	0.379	0.207
	λ	0.980	0.979	0.883	0.767	0.676	0.651	0.505	0.569	0.317

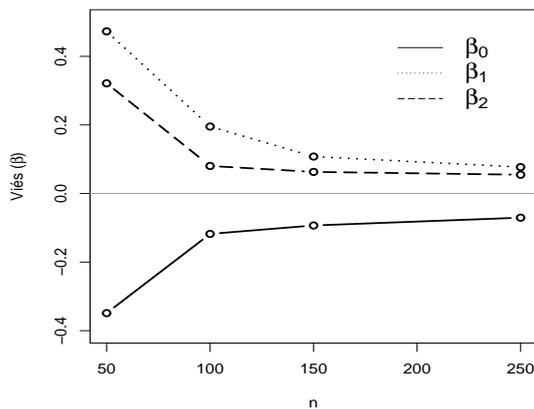
Nas Figuras 4.1 e 4.2 encontram-se os gráficos dos viés relativos a estimativa dos parâmetros das funções de ligações. Na Figura 4.1 temos os gráficos do viés obtidos com as funções probito, logito, complemento log-log, transformação box-cox e probito-assimétrico, no caso da inferência bayesiana sem o uso de variáveis auxiliares. Na Figura 4.2 encontram-se os gráficos do viés das funções probito, logito e probito-assimétrico, no caso da inferência bayesiana com o uso de variáveis auxiliares. Como podemos verificar nas figuras, à medida que aumentamos o tamanho da amostra, o viés tende a zero e o viés para o caso da inferência bayesiana com o uso de variáveis auxiliares em geral é menor.



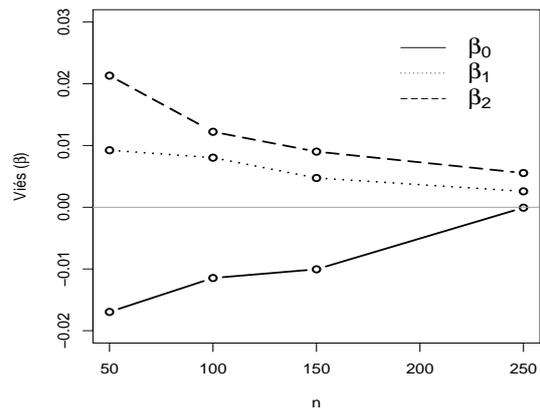
(a) Ligação proibito.



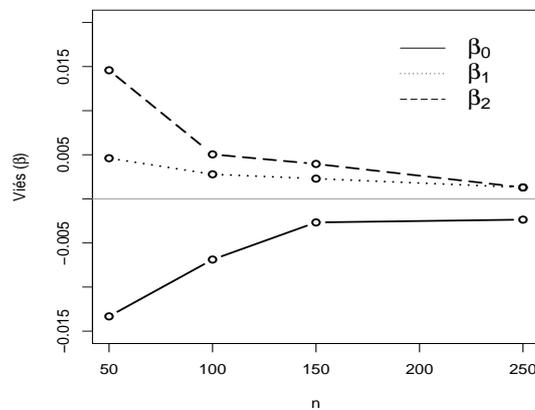
(b) Ligação logito.



(c) Ligação Complemento log-log.

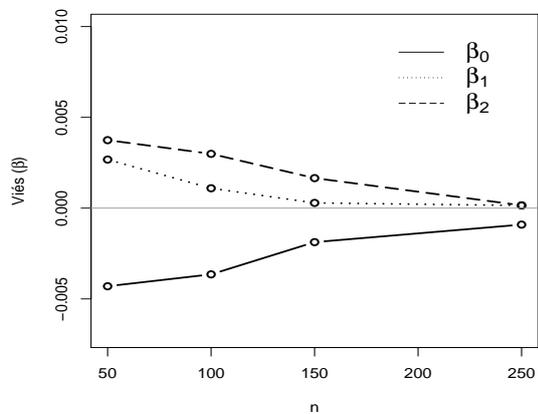


(d) Ligação box-cox.

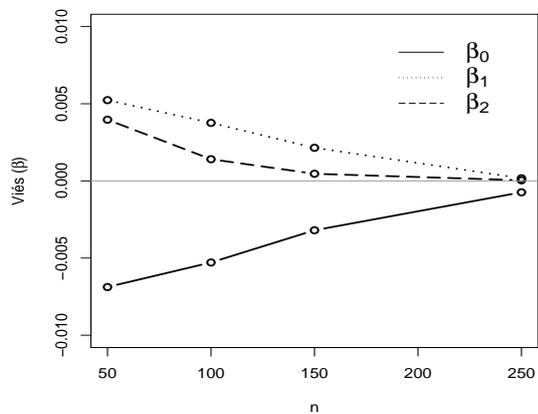


(e) Ligação proibito-assimétrico.

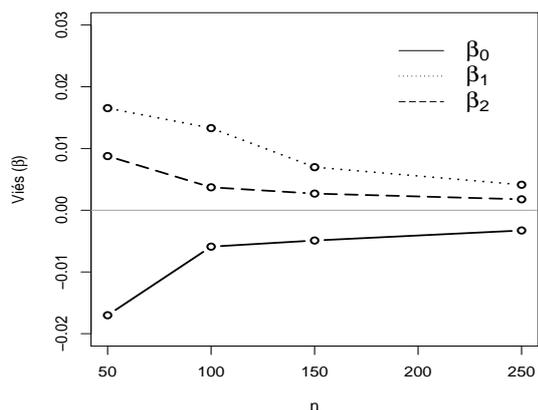
Figura 4.1: Viés encontrado na estimativa dos parâmetros das funções de ligações usando inferência bayesiana sem o uso de variáveis auxiliares.



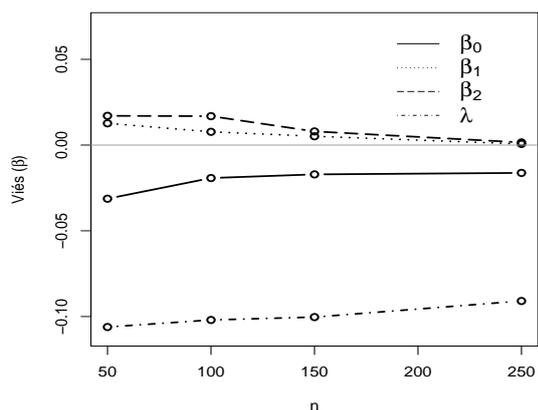
(a) Probito.



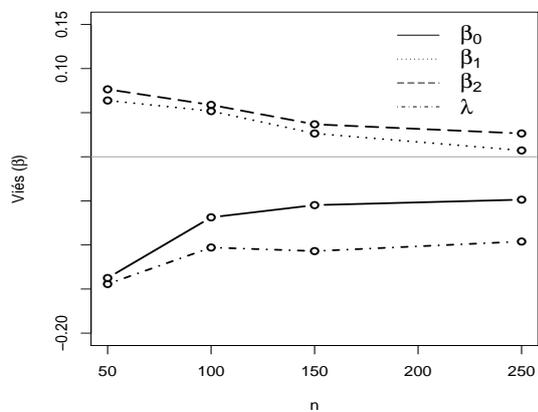
(b) Logito - Algoritmo I.



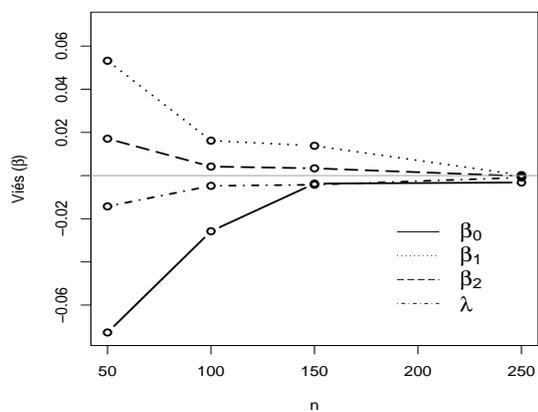
(c) Logito - Algoritmo II.



(d) Probito-assimétrico - Algoritmo I.



(e) Probito-assimétrico - Algoritmo II.



(f) Probito-assimétrico - Algoritmo III.

Figura 4.2: Viés encontrado na estimativa dos parâmetros das funções de ligações usando inferência bayesiana com o uso de variáveis auxiliares.

Aplicações

5.1 Aplicação aos Dados do Ibovespa

Este conjunto de dados se refere a variação do Ibovespa de uma amostra de 883 observações no período de 02/01/13 a 01/08/16 (UolEconomia (2016)). Estabelecemos, se a variação do Ibovespa no dia t com o dia $t - 1$ for positiva atribuímos 1, caso contrário, atribuímos 0. A covariável que escolhemos para explicar essa variação foi o valor do dólar (valor de venda).

Neste conjunto de dados, estamos interessados em saber se existe relação entre a variável do Ibovespa com o valor de fechamento do dólar. Para isso, foi feita uma análise utilizando o modelo de regressão binária, com as funções probito, logito, complemento log-log, transformação box-cox e probito-assimétrico, nas abordagens clássica e bayesiana com uso de variáveis auxiliares. Na abordagem bayesiana optamos pelos algoritmos que apresentaram os melhores resultados nas simulações, logito algoritmo I e o probito-assimétrico algoritmo III.

Os resultados desta análise, abordagens clássica e bayesiana, estão apresentados nas Tabelas 5.1 e 5.2, respectivamente. Na Tabela 5.1 encontram-se a estimativa que é dada pela média da *a posteriori*, o desvio padrão, o valor-z, AIC e BIC para cada modelo. Na Tabela 5.2 encontram-se a estimativa, intervalo de credibilidade (I.C.) de 95%, critério de Geweke e os critérios de seleção de funções de ligações E-AIC e E-BIC.

Tabela 5.1: Resultado da análise dos dados do Ibovespa utilizando inferência clássica.

Funções	Efeitos	Estimativas	DP	Pr(> z)	AIC	BIC
Probito	Constante	-0.031	0.043	0.472	1152.844	1162.411
	Variação do Dólar	-11.821	1.455	4.55e-16		
Logito	Constante	-0.047	0.070	0.502	1151.721	1161.288
	Variação do Dólar	-19.797	2.485	1.66e-15		
Complemento log-log	Constante	-0.424	0.050	< 2e-16	1157.780	1167.347
	Variação do Dólar	-12.262	1.619	3.73e-14		
Transformação box-cox	Constante	-0.003	0.070	0.955	1128.219	1142.569
	Variação do Dólar	-18.653	2.460	3.46e-14		
Probito-assimétrico	Constante	0.522	0.070	1.01e-13	1128.076	1142.426
	Variação do Dólar	-9.760	2.484	8.52e-05		
Probito-assimétrico s/ 210	Constante	0.501	0.070	1.01e-13	1126.465	1141.900
	Variação do Dólar	-10.221	2.532	5.01e-07		

Probito-assimétrico s/210: Probito-assimétrico sem a observação 210.

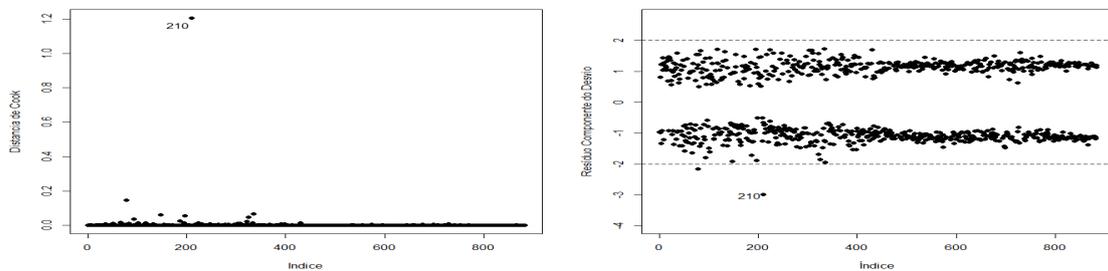
Tabela 5.2: Resultado da análise dos dados do Ibovespa utilizando inferência bayesiana.

Modelos	Efeitos	Estimativas	I.C.(95 %)	Critério de Geweke	E-AIC	E-BIC
Probito	Constante	-0.821	[-2.171; 0.527]	1.197	1789.458	1799.024
	Variação do Dólar	-11.971	[-13.382; -10.560]	-1.441		
Logito	Constante	-0.754	[-2.256; 0.7472]	0.119	1791.437	1801.004
	Variação do Dólar	-15.015	[-16.372; -13.658]	0.524		
Probito-assimétrico	Constante	-1.618	[-3.048; -0.188]	0.464	1782.934	1797.284
	Variação do Dólar	-11.985	[-13.382; -10.587]	-1.085		

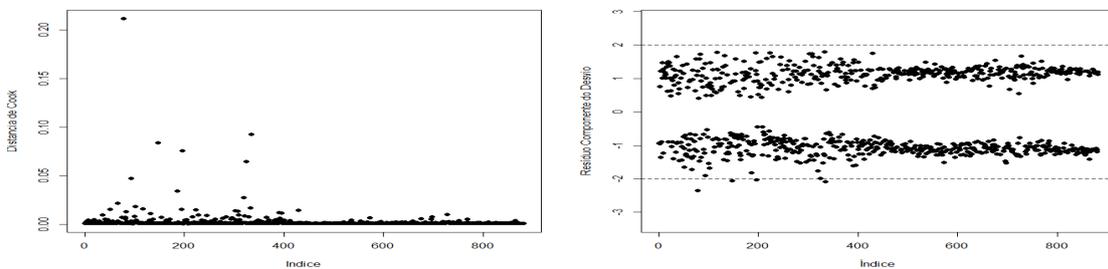
Podemos observar pelas Tabelas 5.1 e 5.2 que o intercepto não é significativo para as funções probito, logito e transformação box-cox. Considerando as funções com intercepto significativo, o modelo que melhor ajustou os dados é probito-assimétrico, segundo os critérios AIC, BIC, E-AIC e E-BIC.

A razão de chances é definida como a razão entre a chance de ocorrer o evento A pela chance de ocorrer o evento B. Sendo, a chance é dada pela probabilidade de ocorrência deste evento dividida pela probabilidade de não ocorrência do mesmo evento. Calculando a razão de chances para este conjunto de dados, concluímos, se houver um aumento de 0.1 na variação do dólar, a chance da variação do Ibovespa ser negativa é maior que 80%.

Na Figura 5.2, apresentamos os gráficos de resíduos dessa função. Na inferência clássica os resíduos apresentados são: distância de cook e resíduo componente do desvio e na inferência bayesiana apresentamos o gráfico de identificação de pontos influentes. Como podemos observar na Figura 5.2 (a), a observação 210 foi apontado como possível ponto influente, por isso, retiramos e fizemos uma nova análise. Os resultados dessa análise encontra-se na Tabela 5.1, denotado por probito-assimétrico sem 210. Como não houve mudança significativa na inferência optamos por não retirá-lo.

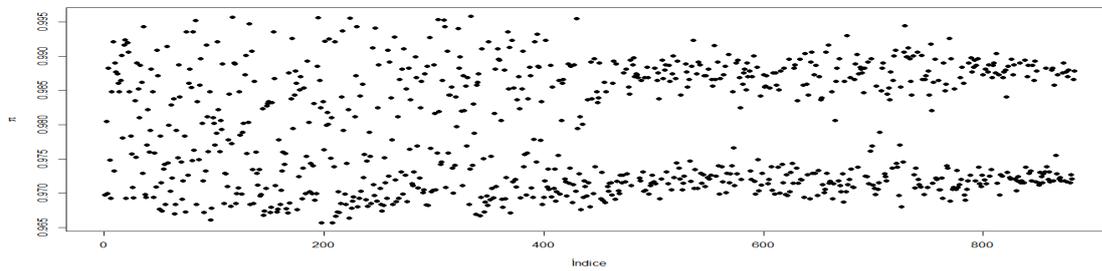


(a) Inferência clássica, com a observação 210.



(b) Inferência clássica, sem a observação 210.

Figura 5.1: Estudo do resíduo usando a função probito-assimétrico na inferência clássica.



(a) Inferência bayesiana, com o uso de variáveis auxiliares.

Figura 5.2: Estudo do resíduo usando a função probito-assimétrico na inferência bayesiana.

5.2 Aplicação aos Dados Educacionais

Este conjunto de dados se refere a uma amostra de tamanho 4174 de um questionário aplicado aos alunos do 3º ano do Ensino Médio das escolas estaduais do estado de São Paulo que participaram da avaliação oficial Prova Brasil do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) (2013).

Neste conjunto de dados, estamos interessados em saber quais são os fatores que influenciam na aprovação do aluno. Para isso, denotamos por 1, se o aluno não reprovou em nenhuma série, e 0, caso contrário. As covariáveis utilizadas são:

Sexo	Etnia	Até que a série a mãe estudou	Tempo ¹	Trabalho ²
Masculino	Branco(a)	Nunca estudou	menos 1 hora	Trabalha
Feminino	Pardo(a)	[0; 4ª série/5º ano[1 hora < t < 2 horas	Não trabalha
	Preto(a)	[4ª série/5º ano; 8ª série/9º ano[2 horas < t < 3 horas	
	Amarelo(a)	[8ª série/9º ano; Ensino Médio[mais 3 horas	
	Indígena	[Ensino Médio; faculdade[Não uso	
	Não sei	[Completo a faculdade[
		Não sei		

¹ Quanto tempo você gasta assistindo à TV, navegando na internet ou jogando.

² Atualmente trabalha fora de casa.

Neste conjunto de dados os modelos foram ajustados com as funções de ligações probito, logito, complemento log-log, transformação box-cox e probito-assimétrico. Os resultados encontram-se nas Tabelas 5.3 e 5.5. Pode-se verificar que a variável tempo não é significativa. Por isso, ajustamos o modelo sem essa variável e apresentamos os resultados nas Tabelas 5.4 e 5.5. Segundo os critérios AIC, BIC, E-AIC e E-BIC, os

melhores ajustes foram obtidos com as funções logito e probito, nas inferências clássica e bayesiana, respectivamente.

Tabela 5.3: Análise de dados educacionais usando inferência clássica.

Funções	Efeitos	Estimativas	DP	Pr($> z $)	AIC	BIC
Probit	Constante	1.188	0.144	$< 2e-16$	3541.700	3579.716
	Sexo	-0.253	0.048	1.36 e-07		
	Etnia	-0.115	0.024	1.85e-06		
	Escolarid. da mãe	0.070	0.017	3.48e-05		
	Tempo	-0.005	0.020	0.801		
	Trabalho	-0.336	0.049	1.32e-11		
Logito	Constante	2.001	0.256	6.11e-15	3541.700	3579.695
	Sexo	-0.465	0.087	1.17e-07		
	Etnia	-0.202	0.042	1.49e-06		
	Escolarid. da mãe	0.127	0.030	2.85e-05		
	Tempo	-0.007	0.036	0.835		
	Trabalho	-0.600	0.088	1.24e-11		
Complemento log-log	Constante	0.773	0.123	3.32e-10	3542.200	3580.241
	Sexo	-0.206	0.039	2.08e-07		
	Etnia	-0.100	0.021	3.44e-06		
	Escolarid. da mãe	0.058	0.014	5.22e-05		
	Tempo	-0.005	0.017	0.759		
	Trabalho	-0.284	0.042	2.25e-11		
Transformação Box-cox	Constante	3.403	0.258	1.06e-03	3541.131	3549.111
	Sexo	-1.029	0.087	9.87e-14		
	Etnia	-0.390	0.042	4.96e-07		
	Escolarid. da mãe	0.257	0.030	4.16e-06		
	Tempo	-0.015	0.036	0.675		
	Se o aluno Trabalha	-1.251	0.089	9.28e-04		
Probit- assimétrico	Constante	1.549	0.256	1.50e-09	3541.643	3579.663
	Sexo	-0.220	0.087	0.012		
	Etnia	-0.099	0.042	0.018		
	Escolarid. da mãe	0.061	0.030	0.045		
	Tempo	-0.004	0.036	0.907		
	Trabalho	-0.289	0.088	0.001		

Tabela 5.4: Análise de dados educacionais usando inferência clássica, sem a variável uso de internet.

Funções	Efeitos	Estimativas	DP	Pr(> z)	AIC	BIC
Probito	Constante	1.170	0.124	< 2e-16	3539.800	3571.444
	Sexo	-0.253	0.048	1.36e-07		
	Etnia	-0.115	0.024	1.86e-06		
	Escolarid. da mãe	0.070	0.017	3.59e-05		
	Trabalho	-0.334	0.049	9.83e-12		
Logito	Constante	1.973	0.220	< 2e-16	3539.700	3571.402
	Sexo	-0.465	0.087	1.16e-07		
	Etnia	-0.202	0.042	1.49e-06		
	Escolarid. da mãe	0.127	0.030	2.90e-05		
	Trabalho	-0.598	0.087	8.71e-12		
Complemento log-log	Constante	0.754	0.106	1.65e-06	3540.300	3572.003
	Sexo	-0.206	0.039	2.10e-07		
	Etnia	0.100	0.021	3.49e-06		
	Escolarid. da mãe	0.058	0.014	5.44e-04		
	Trabalho	-0.282	0.042	1.82e-022		
Transformação box-cox	Constante	3.346	0.258	1.94e-06	3540.215	3547.034
	Sexo	-1.030	0.087	9.31e-07		
	Etnia	-0.389	0.042	6.16e-06		
	Escolarid. da mãe	0.256	0.030	5.65e-04		
	Trabalho	-1.246	0.036	7.66e-05		
Probito- assimétrico	Constante	1.533	0.220	3.45e-12	351.203	3579.323
	Sexo	-0.220	0.087	0.012		
	Etnia	-0.099	0.042	0.018		
	Escolarid. da mãe	0.060	0.030	0.045		
	Trabalho	-0.288	0.087	9.86e-3		

Tabela 5.5: Resultados da análise dos dados educacionais usando inferência bayesiana com o uso de variáveis auxiliares.

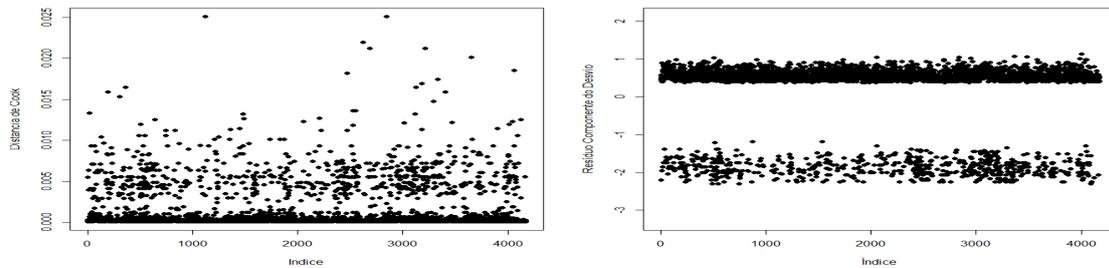
Inferência bayesiana com o uso de variáveis auxiliares						
Modelos	Efeitos	Estimativas	I.C.(95%)	Critério Geweke	E-AIC	E-BIC
Probito	Constante	1.492	[1.041, 1.943]	0.236	4716.530	4720.550
	Sexo	-1.009	[-1.462; -0.555]	-0.624		
	Etnia	-0.683	[-1.130; -0.236]	0.336		
	Escolarid. da mãe	0.764	[0.370; 1.159]	-1.365		
	Tempo	0.023	[-0.458; 0.505]	0.940		
	Trabalho	-0.885	[-1.303, -0.468]	1.189		
Logito	Constante	1.189	[0.743; 1.636]	-1.084	4731.372	4775.729
	Sexo	-1.918	[-2.415; -1.421]	0.101		
	Etnia	-0.763	[-1.206; -0.319]	-1.259		
	Escolarid. da mãe	0.892	[0.566; 1.219]	-1.273		
	Tempo	0.005	[-0.334; 0.324]	1.602		
	Trabalho	-1.275	[-1.710; -0.839]	-0.534		
Probito-assimétrico	Constante	1.478	[1.063; 1.894]	-0.523	4723.003	4731.010
	Sexo	-1.180	[-1.572; -0.788]	0.663		
	Etnia	-0.635	[-1.071; -0.200]	0.129		
	Escolarid. da mãe	1.404	[1.114; 1.694]	-0.248		
	Tempo	0.105	[-0.355; 0.565]	1.266		
	Trabalho	-0.858	[-1.293; -0.424]	-1.604		
Inferência bayesiana com o uso de variáveis auxiliares e sem a variável tempo						
Probito	Constante	1.006	[0.866; 1.146]	-0.462	4712.996	4718.679
	Sexo	-0.257	[-0.394; -0.119]	-0.642		
	Etnia	-0.155	[-0.292; -0.018]	-0.843		
	Escolarid. da mãe	0.067	[0.021; 0.216]	-1.186		
	Trabalho	-0.235	[-0.377; -0.094]	0.199		
	Logito	Constante	1.165	[0.814; 1.516]		
Sexo		-0.878	[-1.243; -0.514]	-0.535		
Etnia		-0.884	[-1.371; -0.397]	0.572		
Escolarid. da mãe		0.624	[0.349; 0.900]	-1.209		
Trabalho		-0.647	[-1.083; -0.212]	-0.437		
Probito-assimétrico		Constante	1.206	[0.786; 1.626]	-0.797	4721.304
	Sexo	-0.720	[-1.159; -0.281]	-0.053		
	Etnia	-0.497	[-0.804; -0.191]	-0.418		
	Escolarid. da mãe	0.585	[0.439; 0.730]	-1.458		
	Trabalho	-0.696	[-1.092; -0.301]	-0.687		

Na Tabela 5.6 apresentamos os resultados obtidos com os cálculos de chances. Nesse caso utilizamos a função logito, nas abordagens clássica e bayesiana e consideramos as seguintes características: masculino e feminino, etnia: branco e negro, trabalho: sim ou não, escolaridade da mãe: nunca estudou e superior completo. Com os resultados obtidos na inferência clássica, calculamos a razão de chances e concluímos: se o aluno do sexo feminino a chance de não reprovar em nenhuma série é 37% menor em relação ao aluno do sexo oposto. Ainda, se o aluno é negro ou trabalha, a chance de não reprovar reduz em 33%. Considerando um aluno negro, masculino e trabalha, ele tem 77% menos chance de não reprovar em relação ao aluno branco, feminino e que trabalha. Além disso, concluímos, se o aluno for do sexo masculino, negro, trabalha, a mãe nunca estudou, a chance de não reprovar reduz em 86% em relação a um aluno de sexo feminino, branco, não trabalha, a mãe tem o ensino superior complemento.

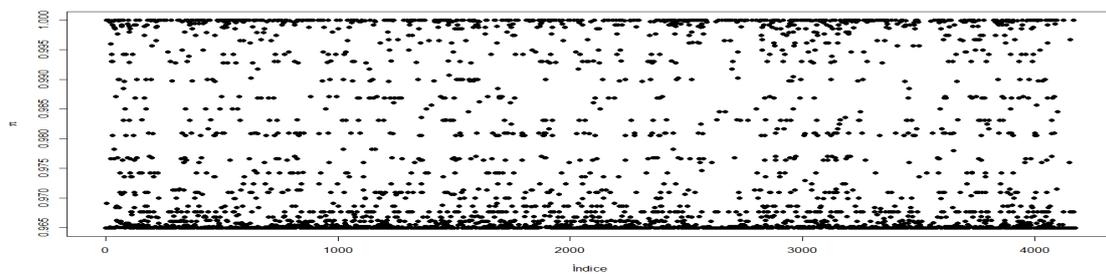
Tabela 5.6: Resultados dos cálculos da razão de chances nas abordagens clássica e bayesiana.

Razão de Chances		
Características	Inferência clássica	Inferência bayesiana
Masculino	4.519	3.408
Feminino	7.198	5.366
Negro	3.197	1.854
Branco	4.798	3.108
Trabalha	3.958	3.536
Não trabalha	7.198	5.366
Nunca estudou	9.288	6.878
Superior completo	15.462	11.613
Negro, Masculino, Trabalha	1.103	0.840
Negro, Masculino, Não Trabalha	2.007	1.225
Negro, Feminino, Trabalha	1.758	1.267
Negro, Feminino, Não Trabalha	3.197	1.854
Branco, Masculino, Trabalha	1.656	2.097
Branco, Masculino, Não Trabalha	3.012	3.108
Branco, Feminino, Trabalha	2.638	2.097
Branco, Feminino, Não Trabalha	4.798	3.108
Mãe: Nunca Estudou, Negro, Masculino, Trabalha	1.424	1.520
Mãe: Curso Superior, Negro, Masculino, Trabalha	2.371	2.357
Mãe: Nunca Estudou, Negro, Feminino, Trabalha	2.268	2.312
Mãe: Curso Superior, Negro, Feminino, Trabalha	3.777	3.642
Mãe: Nunca Estudou Branco, Masculino, Não Trabalha	3.886	2.528
Mãe: Curso Superior, Branco, Masculino, Não Trabalha	6.469	3.999
Mãe: Nunca Estudou, Branco, Feminino, Não Trabalha	6.190	3.919
Mãe: Curso Superior, Branco, Feminino, Não Trabalha	10.305	6.357

Na Figura 5.3 encontram-se os gráficos das análises dos resíduos para as funções logito, na inferência clássica, e probito na inferência bayesiana. Como podemos observar não há possíveis pontos influentes.



(a) Inferência clássica - Logito.



(b) Inferência bayesiana com o uso de variável auxiliar - Probit.

Figura 5.3: Estudos dos resíduos utilizando as funções logito e probito nas inferências clássica e bayesiana, respectivamente.

Considerações Finais

Neste trabalho estudamos o modelo de regressão binária no contexto clássico e bayesiano. Na abordagem bayesiana, trabalhamos com duas metodologias para obter a distribuição *a posteriori*. No primeiro caso, consideramos densidades *a priori* não informativa e utilizamos o algoritmo de Metropolis-Hastings para gerar amostras da distribuição *a posteriori*. No segundo caso, adicionamos variáveis auxiliares na construção do modelo e deste modo, obtemos as distribuições condicionais completas com forma conhecida e que facilitou a implementação do algoritmo Amostrador de Gibbs. Porém, como houve correlações nas cadeias, fizemos atualizações conjuntas (em blocos) das quantidades desconhecidas, sugestão feitas por Holmes e Held (2006).

Na inferência bayesiana com uso de variáveis auxiliares, trabalhamos com as funções probito, logito e probito-assimétrico. Para as funções, probito e logito, apresentamos e implementamos os algoritmos de simulação propostos por Holmes e Held (2006). Na função probito-assimétrico, trabalhamos com a proposta feita por Chen et al. (1999) e implementamos a proposta feita por Farias (2007).

Com os estudos de simulações na abordagem clássica, podemos concluir, o percentual de cobertura do intervalo de confiança assintótico está de acordo com o esperado pela teoria. Em relação ao percentual de escolha, concluímos, os critérios AIC e BIC conseguem escolher a função correta com um percentual de acerto superior a 60%, quando as funções têm o mesmo número de parâmetros.

Na abordagem bayesiana, sem o uso de variáveis auxiliares, concluímos, para amostras pequenas, os resultados foram mais precisos para as funções assimétricas, complemento log-log e probito-assimétrico. Comparando as duas metodologias bayesianas, o algoritmo mais eficiente é com o uso de variáveis auxiliares segundo os critérios EQM, MAPE e SMAPE, além disso, os algoritmos que apresentam melhores resultados foram I e III para as funções logito e probito-assimétrico, respectivamente. Com os resultados do cálculo do viés, concluímos, os estimadores são assintoticamente não viciados para as duas metodologias bayesianas.

Em relação à análise de dados da variação do Ibovespa, tanto na abordagem clássica como bayesiana, o intercepto não é significativo para as funções probito, logito e transformação box-cox. Na inferência clássica, o estudo de resíduo detectou a observação 210 como possível ponto influente. Porém, quando retiramos essa observação, não houve mudança significativa na inferência. Considerando as funções com intercepto significativo, a função probito-assimétrico apresentou o melhor ajuste segundo os critérios, AIC, BIC, E-AIC e E-BIC. Além disso, concluímos, se houver um aumento de 0.1 na variação do dólar, a chance da variação do Ibovespa ser negativa é superior a 80%.

Por fim, com as análises feitas para os dados educacionais, concluímos, o tempo gasto assistindo à TV, navegando na internet ou jogando jogos eletrônicos, fora do horário da aula não são significativas. Segundo os critérios AIC, BIC, E-AIC e E-BIC, as funções probito e logito, ajustaram melhor o modelo nas inferências clássica e bayesiana, respectivamente. Considerando a inferência clássica, concluímos, o aluno do sexo feminino, a chance de não reprovar é 37% menor em relação ao aluno do sexo masculino. Ainda, se o aluno é negro ou trabalha, a chance de não reprovar reduz em 33%. E, se o aluno for do sexo masculino, negro, trabalha e a mãe nunca estudou, a chance de não reprovar em nenhum série reduz em 86% em relação ao aluno do sexo masculino, negro, a mãe nunca estudou e que trabalha.

Como proposta futura podemos destacar:

1. Uso de variáveis auxiliares nos modelos assimétricos:
 - (a) Transformação box-cox
 - (b) Aranda-Ordaz (1981)
 - (c) Stukel (1988)
2. Usar o critério WAIC para selecionar funções de ligações.

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-se 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 i -ésima linha da matriz A pela j -ésima coluna da matriz B . `##` denota-se como linhas comentadas.

A.1 Procedimento para amostragem de Probito

O código para o modelo probito assume *a priori*, $\beta \sim N_p(\mathbf{b}, \mathbf{v})$ e a matriz X com dimensão $(n \times p)$.

Algoritmo 1: ALGORITMO HH

Saída: Amostra *a posteriori* do modelo Probito

```

1  ## Armazena-se as constantes inalteradas de MCMC
2   $V \leftarrow (X^T X + v^{-1})^{-1}$ 
3   $S \leftarrow V X^T$ 
4  para  $j = 1$  até número de observações faça
5  |    $H[j] \leftarrow X[j, ] S[, j]$ 
6  |    $T[j] \leftarrow H[j] / (1 - H[j])$ 
7  |    $Q[j] \leftarrow T[j] + 1$ 
8  fim
9  para  $j = 1$  até número de observações faça
10 |  ## Inicializa-se as variáveis latentes  $Z$ , de normais truncadas independentes
11 |   $Z \sim N_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$ 
12 |   $B \leftarrow V(V^{-1} B + X^T Z)$  ##  $B$  é denota a média condicional de  $\beta$ 
13 |  para  $i = 1$  até número de iterações faça
14 |  |   para  $j = 1$  até número de observações faça
15 |  |   |    $Z_{ant} \leftarrow Z[j]$ 
16 |  |   |    $m \leftarrow X[j, ] B$ 
17 |  |   |    $m \leftarrow m - T[j](Z[j] - m)$ 
18 |  |   fim
19 |  |   ## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
20 |  |    $Z[j] \sim N(m, Q[j]) \text{Ind}(Y[j], Z[j])$ 
21 |  |   ## Amostra-se de  $B$  através da seguinte relação
22 |  |    $B \leftarrow B + (Z[j] - Z_{ant}) S[, j]$ 
23 |  fim
24 |  ## Amostra-se de  $\beta$  a partir da distribuição normal p-variada
25 |   $\beta[, i] \sim N_p(B, V)$ 
26 |  ## Obtém-se os resíduos
27 |  para  $j = 1$  até número de observações faça
28 |  |    $r[i, j] \leftarrow Y[j] - \Phi(X[j, ] \beta[, i])$ 
29 |  |    $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ] \beta[, i]$ 
30 |  fim
31 fim

```

A.2 Procedimentos para amostragem no modelo logístico

Considera-se no código que, $\beta \sim N_p(\mathbf{b}, \mathbf{v})$ *a priori* e a matriz X tem dimensão $(n \times p)$.

Algoritmo 2: ALGORITMO I

Saída: Amostra *a posteriori* da variável de mistura no modelo logístico

```

1 ## Inicializa-se as variáveis de mistura
2  $\Delta \leftarrow \mathbf{I}_n$ 
3 ## Inicializa-se as variáveis latentes  $Z$ , de normais truncadas independentes
4  $Z \sim N_n(\mathbf{0}, \mathbf{I}_n) \text{Ind}(Y, Z)$ 
5 para  $i = 1$  até número de iterações faça
6    $V \leftarrow (X^\top \Delta^{-1} X + v^{-1})^{-1}$ 
7    $S \leftarrow V X^\top$ 
8    $B \leftarrow V(v^{-1}b + X^\top \Delta^{-1} Z)$  ## denota a média condicional de  $\beta$ 
9   para  $j = 1$  até número de observações faça
10     $Z_{ant} \leftarrow Z[j]$ 
11     $H[j] \leftarrow X[j, ] S[, j]$ 
12     $W[j] \leftarrow H[j] / (\Delta[j, j] - H[j])$ 
13     $m \leftarrow X[j, ] B$ 
14     $m \leftarrow m - W[j](Z[j] - m)$ 
15     $q \leftarrow \Delta[j, j](W[j] + 1)$ 
16    ## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
17     $Z[j] \sim N(m, q) \text{Ind}(Y[j], Z[j])$ 
18    ## Amostra-se de  $B$  através da seguinte relação
19     $B \leftarrow B + \left( \frac{Z[j] - Z_{ant}}{\Delta[j, j]} \right) S[, j]$ 
20  fim
21  ## Amostra-se de  $\beta$  a partir da distribuição normal p-variada
22   $\beta[, i] \sim N_p(B, V)$ 
23  ## Amostra-se das variáveis de mistura
24  para  $j = 1$  até número de observações faça
25     $R \leftarrow (Z[j] - X[j, ] \beta[, i])$ 
26     $\Delta[j, j] \sim \pi(\delta | \mathbf{z}, \beta)$ 
27    ## Ver Algoritmo 4
28  fim
29  ## Obtém-se os Resíduos
30  para  $j = 1$  até número de observações faça
31     $r[i, j] \leftarrow Y[j] - F_{Lo}(X[j, ] \beta[, i])$  ##  $F_{Lo}$  denota-se a fda de uma mistura
    de  $Lo(0, 1)$ 
32     $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ] \beta[, i]$ 
33     $\tau[i, j] \leftarrow \epsilon^*[i, j] / \sqrt{\Delta[i, j]}$ 
34  fim
35 fim

```

Algoritmo 3: ALGORITMO II

Saída: Amostra *a posteriori* da variável de mistura no modelo logístico

```

1 ## Inicializa-se as variáveis de mistura
2  $\Delta \leftarrow \mathbf{I}_n$ 
3 ## Inicializa-se as variáveis latentes Z, de normais truncadas independentes
4  $Z \sim Lo_n(0, 1)\text{Ind}(Y, Z)$ 
5 para  $i = 1$  até número de iterações faça
6    $V \leftarrow (X^\top \Delta^{-1} X + v^{-1})^{-1}$ 
7    $B \leftarrow V(v^{-1}b + X^\top \Delta^{-1} Z)$ 
8   ## Amostra-se os valores de  $\beta$  a partir da distribuição normal p-variada
9    $\beta[, i] \sim N_p(B, V)$ 
10  ## Atualiza-se conjuntamente  $\{Z, \Delta\}$ 
11  para  $j = 1$  até número de observações faça
12     $m \leftarrow X[j, ]\beta[, i]$ 
13    ## Amostra-se de  $Z[j]$ , da distribuição logística truncada
14     $Z[j] \sim Lo(m, 1)\text{Ind}(Y[j], Z[j])$ 
15    ## Amostra-se das variáveis de mistura
16     $R \leftarrow (Z[j] - X[j, ]\beta[, i])$ 
17     $\Delta \sim \pi(\delta | \mathbf{z}, \beta)$ 
18    ## Ver Algoritmo 4
19  fim
20  ## Obtém-se os Resíduos
21  para  $j = 1$  até número de observações faça
22     $r[i, j] \leftarrow Y[j] - F_{Lo}(X[j, ]\beta[, i])$  ##  $F_{Lo}$  denota-se a fda de uma mistura
    de  $Lo(0, 1)$ 
23     $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\beta[, i]$ 
24     $\tau[i, j] \leftarrow \epsilon^*[i, j] / \sqrt{\Delta[i, j]}$ 
25  fim
26 fim

```

Algoritmo 4: PROCEDIMENTO PARA AMOSTRAR DE $\delta \sim \pi(\delta|z, \beta)$

```

1 repita
2   ## Temos que  $r^2 = (z_i - \mathbf{x}_i^\top \beta)^2$ 
3   ## Para iniciar, gera-se uma amostra da variável proposta
4    $Y \sim \chi_1^2$ 
5    $Y \leftarrow 1 + (Y - \sqrt{Y(4r + Y)})/(2r)$ 
6    $U \sim U[0, 1]$ 
7   if  $U \leq 1/(1 + Y)$  then
8     |  $\delta \leftarrow r/Y$ 
9     | Senão  $\delta \leftarrow rY$ 
10  end
11  ## Gera-se de  $\delta \sim GIG(0.5; 1, r^2)$ 
12   $U \sim U(0, 1)$ 
13  if  $\delta > 4/3$  then
14    | OK  $\leftarrow direita(U, \delta)$ 
15  end
16  else
17    | OK  $\leftarrow esquerda(U, \delta)$ 
18  end
19  até OK = 1;
20
```

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

Algoritmo 5: O PSEUDO-CÓDIGO PARA A REGIÃO DIREITA

```

1 OK  $\leftarrow$  direita( $U, \delta$ )
2  $Z \leftarrow 1$ 
3  $X \leftarrow \exp(0.5\delta)$ 
4  $j \leftarrow a \leftarrow b \leftarrow 0$ 
5 repita
6    $j \leftarrow j + 1$ 
7    $Z \leftarrow Z - (j + 1)^2 X^{(j+1)^2-1}$ 
8   if ( $Z > U$ ) then
9     OK  $\leftarrow 1$ 
10     $a \leftarrow a + 1$ 
11     $b \leftarrow b - 1$ 
12  end
13   $j \leftarrow j + 1$ 
14   $Z \leftarrow Z + (j + 1)^2 X^{(j+1)^2-1}$ 
15  if  $Z \leq U$  then
16    OK  $\leftarrow 0$ 
17     $a \leftarrow a - 1$ 
18     $b \leftarrow b + 1$ 
19  end
20 até  $a = 2$  ou  $b = 2$ ;
21 Retorne OK.

```

Algoritmo 6: O PSEUDO-CÓDIGO PARA A REGIÃO ESQUERDA

```

1 OK  $\leftarrow$  esquerda( $U, \delta$ )
2  $H \leftarrow 0.5 \log(2) + 2.5 \log(\pi) - 2.5 \log(\delta) - \frac{\pi}{2\delta} + 0.5\delta$ 
3  $lU \leftarrow \log(U)$ 
4  $Z \leftarrow 1$ 
5  $K \leftarrow \delta/\pi$ 
6  $X \leftarrow \exp(-\pi^2/(2\delta))$ 
7  $j \leftarrow a \leftarrow b \leftarrow 0$ 
8 repita
9    $j \leftarrow j + 1$ 
10   $Z \leftarrow Z - KX^{j^2-1}$ 
11  if  $H + \log(Z) > lU$  then
12    OK  $\leftarrow 1$ 
13     $a \leftarrow a + 1$ 
14     $b \leftarrow b - 1$ 
15  end
16   $j \leftarrow j + 1$ 
17   $Z \leftarrow Z + (j + 1)^2 X^{(j+1)^2-1}$ 
18  if  $H + \log(Z) \leq lU$  then
19    OK  $\leftarrow 0$ 
20     $a \leftarrow a - 1$ 
21     $b \leftarrow b + 1$ 
22  end
23  até  $a = 2$  ou  $b = 2$ ;
24  Retorne OK.
25
```

A.3 Procedimentos para amostragem no modelo probito-assimétrico

Algoritmo 7: ALGORITMO HH(\mathbf{z} , $\boldsymbol{\beta}$)

Saída: Amostra *a posteriori* do modelo Probito-Assimétrico

```

1 ## Armazena-se as constantes inalteradas de MCMC
2  $V \leftarrow (X^\top X + v^{-1})^{-1}$ 
3  $S \leftarrow VX^\top$ 
4 para  $j = 1$  até número de observações faça
5   |  $H[j] \leftarrow X[j, ]S[, j]$ 
6   |  $T[j] \leftarrow H[j]/(1 - H[j])$ 
7   |  $Q[j] \leftarrow T[j] + 1$ 
8 fim
9 ## Inicializa-se as variáveis latentes  $Z$  e  $W$ , de normais truncadas independentes
10  $Z \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(Y, Z)$ 
11  $W \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(W > \mathbf{0})$ 
12 para  $i = 1$  até número de interações faça
13   |  $B \leftarrow V [V^{-1}b + X^\top (Z - \lambda[i]W)]$  ##  $B$  é denota a média condicional de  $\boldsymbol{\beta}$ 
14   | para  $j = 1$  até número de observações faça
15     |  $Z_{ant} \leftarrow Z[j]$ 
16     |  $m \leftarrow X[j, ]B - \lambda[i]W[j]$ 
17     |  $m \leftarrow m - T[j](Z[j] - m)$ 
18     | ## Amostra-se de  $Z[j]$  a partir da distribuição normal truncada
19     |  $Z[j] \sim N(m, Q[j])\text{Ind}(Y[j], Z[j])$ 
20     | ## Amostra-se de  $B$  através da seguinte relação
21     |  $B \leftarrow B + (Z[j] - Z_{ant})S[, j]$ 
22   | fim
23   | ## Amostra-se de  $\boldsymbol{\beta}$  a partir da distribuição normal p-variada
24   |  $\boldsymbol{\beta}[, i] \sim N_p(B, V)$ 
25   |  $s \leftarrow 1/(1 + \lambda[i]^2)$ 
26   | para  $j = 1$  até número de observações faça
27     |  $r \leftarrow s\lambda[i](Z[j] - X[j, ]\boldsymbol{\beta}[, i], \lambda[i])$ 
28     | ## Amostra-se de  $W$  a partir da distribuição normal truncadas
29     |  $W[j] \sim N(r, s)\text{Ind}(W[j] > \mathbf{0})$ 
30   | fim
31   |  $q \leftarrow 1/(\tau^{-1} + W^\top W)$ 
32   |  $d \leftarrow q [\tau^{-1}\alpha - W^\top (Z - X\boldsymbol{\beta}[, i])]$ 
33   | ## Amostra-se de  $\lambda$  a partir da distribuição normal truncada
34   |  $\lambda[i] \sim N(d, q)$ 
35   | ## Obtém-se os resíduos
36   | para  $j = 1$  até número de observações faça
37     |  $r[i, j] \leftarrow Y[j] - \Phi_{CDS}(X[j, ]\boldsymbol{\beta}[, i], \lambda[i])$ 
38     |  $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\boldsymbol{\beta}[, i]$ 
39     |  $\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$ 
40     |  $\varsigma[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$ 
41   | fim
42 fim

```

Algoritmo 8: ALGORITMO HH(\mathbf{z} , \mathbf{w})

Saída: Amostra *a posteriori* do modelo Probito-Assimétrico

```

1 ## Armazena-se as constantes inalteradas de MCMC
2  $V \leftarrow (X^\top X + v^{-1})^{-1}$ 
3  $S \leftarrow VX^\top$ 
4 para  $j = 1$  até número de observações faça
5   |  $H[j] \leftarrow X[j, ]S[, j]$ 
6   |  $T[j] \leftarrow H[j]/(1 - H[j])$ 
7   |  $Q[j] \leftarrow T[j] + 1$ 
8 fim
9 ## Inicializa-se as variáveis latentes Z e W, de normais truncadas independentes
10  $Z \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(Y, Z)$ 
11  $W \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(W > \mathbf{0})$ 
12 para  $i = 1$  até número de interações faça
13   |  $B \leftarrow V [(v^{-1}b + X^\top(Z - \lambda[i]W))] \quad \#\#$  Denota a média condicional de  $\beta$ 
14   |  $\#\#$  Amostra-se de  $\beta$  a partir da distribuição normal p-variada
15   |  $\beta[, i] \sim N_p(B, V)$ 
16   |  $\#\#$  Atualiza-se conjuntamente  $\{Z, W\}$ 
17   | para  $j = 1$  até número de observações faça
18   |   |  $m \leftarrow X[j, ]\beta[, i] - \lambda[i]W[j]$ 
19   |   |  $\#\#$  Amostra-se de  $Z[j]$  a partir da distribuição normal-assimétrica
20   |   | truncada
21   |   |  $Z[j] \sim SN(m, 1 + \lambda[i]^2, -\lambda[i])\text{Ind}(Y[j], Z[j])$ 
22   | fim
23   |  $s \leftarrow 1/(1 + \lambda[i]^2)$ 
24   | para  $j = 1$  até número de observações faça
25   |   |  $r \leftarrow s\lambda[i](Z[j] - X[j, ]\beta[, i])$ 
26   |   |  $\#\#$  Amostra-se de W a partir da distribuição normal truncadas
27   |   |  $W[j] \sim N(r, s)\text{Ind}(W[j] > \mathbf{0})$ 
28   | fim
29   |  $q \leftarrow 1/(\tau^{-1} + W^\top W)$ 
30   |  $d \leftarrow q [\tau^{-1}\alpha - W^\top(Z - X\beta[, i])]$ 
31   |  $\#\#$  Amostra-se de  $\lambda$  a partir da distribuição normal truncada
32   |  $\lambda[i] \sim N(d, q)$ 
33   |  $\#\#$  Obtém-se os resíduos
34   | para  $j = 1$  até número de observações faça
35   |   |  $r[i, j] \leftarrow Y[j] - \Phi_{CDS}(X[j, ]\beta[, i], \lambda[i])$ 
36   |   |  $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\beta[, i]$ 
37   |   |  $\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$ 
38   |   |  $\varsigma[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$ 
39 fim

```

Algoritmo 9: ALGORITMO HH($\mathbf{z}, \boldsymbol{\beta}, \lambda$)

```

1  ## Inicializa-se as variáveis latentes Z e W, de normais truncadas independentes
2   $Z \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(Y, Z)$ 
3   $W \sim N_n(\mathbf{0}, \mathbf{I}_n)\text{Ind}(W > \mathbf{0})$ 
4  ## Concatena a matriz X e o vetor W
5   $A \leftarrow [X: -W]$ 
6  para  $i = 1$  até número de interações faça
7       $V \leftarrow (A^\top A + v^{-1})^{-1}$ 
8       $S \leftarrow VA^\top$ 
9       $B \leftarrow V(v^{-1}b + A^\top Z)$  ## B denota a média condicional de  $\boldsymbol{\theta}$ 
10     para  $j = 1$  até número de observações faça
11          $H[j] \leftarrow A[j, ]S[, j]$ 
12          $T[j] \leftarrow H[j]/(1 - H[j])$ 
13          $Q[j] \leftarrow T[j] + 1$ 
14          $Z_{ant} \leftarrow Z[j]$ 
15          $m \leftarrow A[j, ]B - T[j](Z[j] - A[j, ]B)$ 
16         ## Amostra-se de Z[j] a partir da distribuição normal truncada
17          $Z[j] \sim N(m, Q[j])\text{Ind}(Y[j], Z[j])$ 
18         ## Atualiza-se B através da seguinte relação
19          $B \leftarrow B + (Z[j] - Z_{ant})S[, j]$ 
20     fim
21     ## Amostra-se de  $\boldsymbol{\theta}$  a partir da distribuição normal  $(n + 1)$  variada
22      $\boldsymbol{\theta}[, i] \sim N_{p+1}(B, V)$ 
23      $\boldsymbol{\beta} \leftarrow \boldsymbol{\theta}[1 : p, i]$ 
24      $\lambda \leftarrow \boldsymbol{\theta}[p + 1, i]$ 
25      $s \leftarrow 1/(1 + \lambda^2)$ 
26     para  $j = 1$  até número de observações faça
27          $r \leftarrow s\lambda(Z[j] - X[j, ]\boldsymbol{\beta})$ 
28         ## Amostra-se de W a partir da distribuição normal truncadas
29          $W[j] \sim N(r, s)\text{Ind}(W[j] > \mathbf{0})$ 
30     fim
31      $q \leftarrow 1/(\tau^{-1} + W^\top W)$ 
32      $d \leftarrow q[\tau^{-1}\alpha - W^\top(Z - X\boldsymbol{\beta})]$ 
33     ## Amostra-se de  $\lambda$  a partir da distribuição normal
34      $\lambda[i] \sim N(d, q)$ 
35     ## Obtém-se os resíduos
36     para  $j = 1$  até número de observações faça
37          $r[i, j] \leftarrow Y[j] - \Phi_{CDS}(X[j, ]\boldsymbol{\beta}[, i], \lambda[i])$ 
38          $\epsilon^*[i, j] \leftarrow Z[j] - X[j, ]\boldsymbol{\beta}[, i]$ 
39          $\epsilon[i, j] \leftarrow \epsilon^*[i, j] - \lambda[i]W[j]$ 
40          $\varsigma[i, j] \leftarrow \epsilon^*[i, j]^2/(1 + \lambda[i]^2)$ 
41     fim
42 fim

```

Método para amostrador da variável misturadora no modelo logístico

Neste apêndice descrevemos como gerar valores da distribuição a posteriori do vetor de variáveis misturadores $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_n\}$ no modelo de regressão logística apresentado na seção 3.2.2.

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

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

sendo $r^2 = (\mathbf{z}_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$. Como esta distribuição não tem uma forma conhecida, usaremos o algoritmo de rejeição para gerar amostra dela.

A distribuição B.1 tem o núcleo parecido com a densidade de uma gaussiana-inversa generalizada, denotada por GIG, veja o Apêndice D. Holmes e Held (2006) sugerem o uso da GIG com os parâmetros $\lambda = 0.5, \psi = 1, \chi = r^2$, como distribuição proposta para implementação do algoritmo de rejeição.

Utilizando as propriedades do Apêndice D podemos escrever a distribuição gaussiana-inversa generalizada como gaussiana inversa da seguinte maneira

$$GIG(0.5; 1; r^2) = rGIG(0.5; r; r) = \frac{r}{GIG(-0.5; r; r)} = \frac{r}{GI(1; r)}, \quad (\text{B.2})$$

sendo GI denotada por gaussiana inversa, veja no Apêndice D. Usando a relação B.2, podemos amostrar de uma distribuição GIG através de uma distribuição GI, utilizando o algoritmo de inversa Devroye (1986). O algoritmo encontra-se no Apêndice A.

Seja $g(\cdot)$ densidade de probabilidade proposta, denotada por $GIG(0.5; 1; r^2)$, sendo $r^2 = (z_i - x_i^\top \boldsymbol{\beta})^2$. No algoritmo de rejeição, gera-se um valor de $g(\cdot)$, e este valor é aceito para $\pi(\cdot | \mathbf{z}, \boldsymbol{\beta})$ com probabilidade $\alpha(\cdot)$,

$$\alpha(\delta) = \frac{\pi(\delta | \mathbf{z}, \boldsymbol{\beta})}{Mg(\delta)}, \quad (\text{B.3})$$

sendo $M \geq \sup_{\delta} \frac{\pi(\delta | \mathbf{z}, \boldsymbol{\beta})}{g(\delta)}$.

Como a distribuição *a posteriori* de δ é proporcional ao produto de sua distribuição *a priori* pela verossimilhança, temos que B.8 que pode ser reescrita da seguinte forma,

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

sendo $\ell(\delta)$ a verossimilhança $\ell(\delta) = \delta^{-1/2} \exp(-0.5r^2/\delta)$, K é maior ou igual ao $\sup_{\delta} \frac{\ell(\delta)\pi(\delta)}{g(\delta)}$, e $\pi(\delta)$ é a distribuição *a priori* de δ dada pela transformação de variáveis, $\delta_i = (2\psi)^2$, apresentada em 3.8. Onde temos

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

em que $KS(\cdot)$ denota a densidade da distribuição Kolmogorov-Smirnov assintótica dada no Apêndice D.

A distribuição *a priori* dada por B.5 pode ser reescrita em duas formas, sendo elas

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

e

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

Assumindo $K = 1$ em B.8, a probabilidade de aceitação é dada por

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

Como a probabilidade dada em B.8 é uma série infinita, podemos escrever as equações B.6 e B.7, denotadas por densidades A e B em forma de séries alternadas Devroye (1986), pag. 161-165, definida por

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

Reescrevendo a densidade A em função $ch(\cdot)$ e $a_n(\cdot)$ obtemos

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

em que

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

Enquanto que, a densidade B é dada por

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

sendo

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

Para aplicação do método da séries alternadas, devemos verificar se a_n^* satisfazem a condição de monotonicidade. Isto é feito no Lema B.0.1.

Lema B.0.1. O termo a_n na primeira expansão em séries são monótona \downarrow para $\delta > 4/3$. Para a segunda expansão em séries, são monótona \downarrow para $\delta \in (0, \pi^2)$, mas com restrição que $n \geq 0$.

Prova do Lema B.0.1 Para a expansão em série da densidade A, temos

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

Para a expansão em série da densidade B, consideramos n par. Neste caso temos

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

e

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

Verificamos que a última equação não é maior que $2n - 2 \log(n+1) \geq 0$.

Portanto, $\pi(\delta)$ é monótona em $(4/3, \infty)$ e em $(0, \pi^2)$. Note que o intervalo de convergência entre tais séries apresenta um intersecção, $[4/3, \pi^2]$. Usaremos o ponto de quebra $4/3$, pois a série obtida através da densidade B converge mais rapidamente.

Provas das distribuições da função Probito-assimétrico

Lema C.0.1. *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*

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

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

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

Lema C.0.2. *Sejam, X e v : matrizes de dimensões $n \times p$ e $p \times p$, respectivamente; z , w , β e b : vetores de dimensões $n \times 1$, $n \times 1$ e $p \times 1$, respectivamente, e os valores λ , α e τ . Se todas as inversas necessárias existem, então*

$$1. (z - X\beta + \lambda w)^\top (z - X\beta + \lambda w) + (\beta - b)^\top v^{-1}(\beta - b) = (\beta - B)^\top v^{-1}(\beta - B) + C, \text{ sendo}$$

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

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

$$C = (1 + \lambda^2)^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})$$

$$m = -\frac{\lambda}{1 + \lambda^2}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})$$

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

$$C = [\mathbf{z} - (\mathbf{X}\boldsymbol{\beta} - \alpha\mathbf{w})]^\top (\mathbf{I}_n + \tau\mathbf{w}\mathbf{w}^\top)^{-1} [\mathbf{z} - (\mathbf{X}\mathbf{b} - \alpha\mathbf{w})]$$

$$m = v [\tau^{-1}\alpha - \mathbf{w}^\top(\mathbf{z} + \mathbf{X}\boldsymbol{\beta})]$$

$$v = (\tau^{-1} + \mathbf{w}\mathbf{w}^\top)^{-1}$$

Lema C.0.3. *Sejam, \mathbf{X} e \mathbf{v} : matrizes de dimensões $n \times p$, $p \times p$, respectivamente; \mathbf{z} , $\boldsymbol{\omega}$, $\boldsymbol{\beta}$ 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*

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

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

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

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

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

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

Distribuição 3.16: Utilizando o teorema de Bayes, temos que a distribuição condicional completa de $\boldsymbol{\beta}$ é dada por

$$\pi(\boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\omega}, \lambda) = C^{-1} \pi(\mathbf{z}, \boldsymbol{\beta} | \boldsymbol{\omega}, \lambda), \text{ sendo } C = \pi(\mathbf{z} | \boldsymbol{\omega}, \lambda). \quad (\text{C.2})$$

Como as distribuições *a priori* de $\boldsymbol{\beta}$, $\boldsymbol{\omega}$ e λ são independentes sob o modelo 3.15, podemos escrever $\pi(\mathbf{z}, \boldsymbol{\beta} | \boldsymbol{\omega}, \lambda) = \pi(\mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) \pi(\boldsymbol{\beta})$, onde $\pi(\mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda)$ em que,

$$\pi(\mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) = \phi_n(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \lambda\boldsymbol{\omega}, \mathbf{I}_n), \quad (\text{C.3})$$

sendo $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ uma fdp de uma distribuição normal n-variada com o vetor de médias $\boldsymbol{\mu}$ e a matriz de covariâncias $\boldsymbol{\Sigma}$. Assim, temos que

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

Utilizando o lema C.0.2, temos que

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

sendo

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

Substituindo (C.4) em (C.2), temos que

$$\pi(\boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\omega}, \lambda) = C^{-1} \pi(\mathbf{z}, \boldsymbol{\beta} | \boldsymbol{\omega}, \lambda) = C^{-1} K \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{B})^\top \mathbf{V}^{-1} (\boldsymbol{\beta} - \mathbf{B}) \right\} \quad (\text{C.5})$$

Como C e K, dados em (C.2) e (C.4), respectivamente, não dependem de $\boldsymbol{\beta}$ é o fator restante da Equação (C.5) é o núcleo de uma distribuição normal p-variada com o vetor de média \mathbf{B} e a matriz de covariâncias \mathbf{V} . Logo, a função (C.5) só será uma fdp se $C^{-1}K = (2\pi)^{-n/2} |\mathbf{V}|^{1/2}$, o que implica que $C = \pi(\mathbf{z} | \boldsymbol{\omega}, \lambda)$ é a fdp de uma distribuição normal n-variada dado por

$$\pi(\mathbf{z} | \boldsymbol{\omega}, \lambda) = \phi_n(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \lambda\boldsymbol{\omega}, \mathbf{I}_n + \mathbf{X}\mathbf{v}\mathbf{X}^\top) \quad (\text{C.6})$$

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

$$\boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\omega}, \lambda \sim N_p(\mathbf{B}, \mathbf{V})$$

\mathbf{B} e \mathbf{V} são dados em (C.4).

Distribuição 3.17: Utilizando o teorema de Bayes, temos que a distribuição condicional completa de \mathbf{z} é dada por

$$\pi(\mathbf{z} | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) = C^{-1} \pi(\mathbf{y}, \mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda), \text{ sendo } \pi(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda). \quad (\text{C.7})$$

Pode-se escrever, $\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) = \pi(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\omega}, \lambda)\pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda)$, em que $\pi(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\omega}, \lambda)$ é dada em (C.3) e

$$\pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) = \pi(\mathbf{y}|\mathbf{z}) = \text{Ind}(\mathbf{y}, \mathbf{z}) \prod_{i=1}^n \text{Ind}(y_i, z_i) \quad (\text{C.8})$$

onde $\text{Ind}(y_i, z_i) = I(z_i > 0)I(y_i = 1) + I(z_i \leq 0)I(y_i = 0)$, sendo $I(\cdot)$ a função indicadora. Substituindo (C.8) e (C.3) em (C.7), temos que

$$\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\omega}, \lambda) = C^{-1} \phi_n(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \lambda\boldsymbol{\omega}, \mathbf{I}_n) \text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.9})$$

Como $C = \pi(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\omega}, \lambda)$ não depende de \mathbf{z} e o fator restante da distribuição (C.9) é o núcleo de uma distribuição normal n-variada truncada com vetor de médias $\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda\omega$ e matriz de covariâncias \mathbf{I}_n . Portanto, os componentes z_i , $i = 1, \dots, n$ são variáveis aleatórias independentes com distribuição truncada dada por

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

Distribuição 3.18: Utilizando o Teorema de Bayes, temos que a distribuição condicional completa de $\boldsymbol{\omega}$ é dada por

$$\pi(\boldsymbol{\omega}|\mathbf{z}, \boldsymbol{\beta}, \lambda) = C^{-1} \pi(\mathbf{z}, \boldsymbol{\omega}|\boldsymbol{\beta}, \lambda), \text{ sendo } C = \pi(\mathbf{z}|\boldsymbol{\beta}, \lambda). \quad (\text{C.10})$$

Pela independência entre as distribuições *a priori* de $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, e λ , tem-se

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

Utilizando o Lema C.0.2, tem-se

$$\begin{aligned} \pi(\mathbf{z}, \boldsymbol{\omega}|\boldsymbol{\beta}, \lambda) &= K \exp \left\{ -\frac{1 + \lambda^2}{2} (\boldsymbol{\omega} - \mathbf{m})^\top (\boldsymbol{\omega} - \mathbf{m}) \right\} I(\boldsymbol{\omega} > 0), \\ K &= \pi^{-n} \exp \left\{ -\frac{1}{2(1 + \lambda^2)} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ m &= -\frac{\lambda}{1 + \lambda^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

Por fim, substituindo (C.11) em (C.10), tem-se

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

Como C e K não dependem de $\boldsymbol{\omega}$, e o fator restante da distribuição (C.11) é o núcleo de uma distribuição p-variada positiva. Portanto,

$$\boldsymbol{\omega}|\mathbf{z}, \boldsymbol{\beta}, \lambda \propto N_n \left(-\frac{\lambda}{1+\lambda^2}(\mathbf{z} - \mathbf{X}^\top \boldsymbol{\beta}), (1+\lambda^2)^{-1}I_n \right) I(\boldsymbol{\omega} > 0) \quad (\text{C.12})$$

Como a matriz $(1+\lambda^2)^{-1}I_n$ é diagonal, os componentes ω_i, i, \dots, n são variáveis aleatórias independentes com distribuição normal truncada dada por

$$\omega|z_i, \boldsymbol{\beta}, \lambda \propto N \left(-\frac{\lambda}{1+\lambda^2}(z_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \frac{1}{1+\lambda^2} \right) I(\omega_i > 0) \quad (\text{C.13})$$

Distribuição 3.19: Utilizando o teorema de Bayes, temos que a distribuição condicional completa de λ é dada por

$$\pi(\lambda|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega}) = C^{-1}\pi(\mathbf{z}, \lambda|\boldsymbol{\beta}, \boldsymbol{\omega}), \text{ sendo } C^{-1} = \pi(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\omega}). \quad (\text{C.14})$$

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

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

Usando o Lema C.0.2, segue que

$$\pi(\mathbf{z}, \lambda|\boldsymbol{\beta}, \boldsymbol{\omega}) = K \exp \left\{ -\frac{1}{2} \frac{(\lambda - m)^2}{\tau} \right\} \quad (\text{C.15})$$

sendo que,

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

Substituindo (C.15) em (C.14), tem-se

$$\pi(\lambda|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega}) = C^{-1}K \exp \left\{ -\frac{1}{2v}(\lambda - m)^2 \right\} \quad (\text{C.16})$$

Como C e K não dependem de λ e o fator restante da distribuição (C.16) é o núcleo de uma distribuição normal com média m e variância v . Logo, a função (C.16) só será uma fdp se $C^{-1}K = (2\pi v)^{-1/2}$, o que implica $C = \pi(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\omega})$ é uma fdp de uma distribuição normal n -variada dada por

$$\pi(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\omega}) = \phi_n(\mathbf{z}; \mathbf{X}\boldsymbol{\beta} - \alpha\boldsymbol{\omega}, \mathbf{I}_n + \tau\boldsymbol{\omega}^\top\boldsymbol{\omega}), \quad (\text{C.17})$$

é a distribuição condicional completa de λ é dada por:

$$\lambda|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\omega} \sim N(m, v)$$

em que, m e v são dados em C.15.

Proposição C.0.4. *No modelo probito-assimétrico, a distribuição $\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda$ é dada pela expressão (3.20).*

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

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda) = C^{-1}\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\omega}, \lambda), \quad (\text{C.18})$$

sendo $C = \pi(\mathbf{y}|\boldsymbol{\omega}, \lambda)$. Sabendo que $\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\omega}, \lambda) = \pi(\mathbf{z}|\boldsymbol{\omega}, \lambda)\pi(\mathbf{y}|\mathbf{z})$, onde $\pi(\mathbf{z}|\boldsymbol{\omega}, \lambda)$ e $\pi(\mathbf{y}|\mathbf{z})$ são dadas em (C.6) e (C.8), respectivamente. Logo, a distribuição (C.18) pode ser escrita como

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda) = C^{-1}\phi_n(\mathbf{z}; \mathbf{X}\mathbf{b} - \lambda\boldsymbol{\omega}, \mathbf{I}_n + \mathbf{X}\mathbf{v}\mathbf{X}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}). \quad (\text{C.19})$$

A Equação (C.19) só será uma fdp se $C = \bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \mathbf{X}\mathbf{b} - \lambda\boldsymbol{\omega}, \mathbf{I}_n - \mathbf{X}\mathbf{v}\mathbf{X}^\top)$, em que $\bar{\Phi}_n(R(\mathbf{y}, \mathbf{z}); \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a função de distribuição acumulada da normal n -variada na região $R(\mathbf{y}, \mathbf{z}) = \mathbf{z} = (z_1, \dots, z_n)^\top$; $z_i > 0$ se $y_i = 1$ ou $z_i \leq 0$ se $y_i = 0$, com vetor de média $\boldsymbol{\mu}$ e matriz de variâncias $\boldsymbol{\Sigma}$. Logo, a distribuição de \mathbf{z} condicionada à $\{\mathbf{y}, \boldsymbol{\omega}, \lambda\}$ é dada por

$$\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda \propto N_n(\mathbf{X}\mathbf{b} - \lambda\boldsymbol{\omega}, \mathbf{I}_n + \mathbf{X}\mathbf{v}\mathbf{X}^\top) \text{Ind}(\mathbf{y}, \mathbf{z}), \quad (\text{C.20})$$

sendo $\text{Ind}(\mathbf{y}, \mathbf{z})$ a função indicadora. Quando $\lambda = 0$, a equação acima se reduz à distribuição matriz de \mathbf{z} a posteriori do modelo probito. ■

Proposição C.0.5. *Sob o modelo probito-assimétrico, temos que a distribuição $\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \lambda$ é dada por (3.22).*

Prova: Utilizando o teorema de Bayes, a distribuição de $\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \lambda)$ é dada por:

$$\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\beta}, \lambda) = C^{-1}\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \lambda), \text{ em que, } C = \pi(\mathbf{y}|\boldsymbol{\beta}, \lambda). \quad (\text{C.21})$$

Considerando o modelo 3.15, temos que $z_i|\omega_i, \boldsymbol{\beta}, \lambda \sim N(\mathbf{x}_i^\top \boldsymbol{\beta} - \lambda\omega_i, 1)$ e $\omega_i \sim HN(0, 1)$. Utilizando as propriedades das distribuição normal-assimétrica, segue que $\mathbf{z}_i|\boldsymbol{\beta}, \lambda \sim SN(\mathbf{x}_i^\top \boldsymbol{\beta}, 1 + \lambda^2, -\lambda)$. Logo, a distribuição de $\pi(\mathbf{z}|\boldsymbol{\beta}, \lambda)$ é dada por

$$\pi(\mathbf{z}|\boldsymbol{\beta}, \lambda) = \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \boldsymbol{\beta})\right]. \quad (\text{C.22})$$

Sabendo que, $\pi(\mathbf{z}, \mathbf{y}|\boldsymbol{\beta}, \lambda) = \pi(\mathbf{z}|\boldsymbol{\beta}, \lambda)\pi(\mathbf{y}|\mathbf{z})$, sendo $\pi(\mathbf{z}|\boldsymbol{\beta}, \lambda)$ e $\pi(\mathbf{y}|\mathbf{z})$ são dadas em (C.22) e (C.8), respectivamente. Logo,

$$\pi(\mathbf{z}|\boldsymbol{\beta}, \lambda) = \prod_{i=1}^n \phi\left(\frac{z_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sqrt{1 + \lambda^2}}\right) \Phi\left[\frac{\lambda}{\sqrt{1 + \lambda^2}}(z_i - \mathbf{x}_i^\top \boldsymbol{\beta})\right] \prod_{i=1}^n \text{Ind}(y_i, z_i). \quad (\text{C.23})$$

Substituindo (C.23) em (C.21), temos

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

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

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

■

Proposição C.0.6. *Sob o modelo probito-assimétrico, temos que a distribuição $\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}$ é dada por (3.25).*

Observe que, dado $\boldsymbol{\omega}$, o modelo 3.15 se reduz ao modelo probito. Logo, a prova da Proposição C.0.6 é obtida de modo similar à prova da Proposição C.0.4, assumindo $\lambda = 0$.

Proposição C.0.7. *Sob o modelo probito-assimétrico, temos que as distribuições condicionais completas de $(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega}, \lambda)$ são dadas por (3.21).*

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

$$\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\omega}, \lambda) = C^{-1}\pi(\mathbf{z}|\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\omega}, \lambda),$$

sendo $C = \pi(\mathbf{z}_{-i}|\mathbf{y}, \boldsymbol{\omega}, \lambda)$ e \mathbf{z}_{-i} denota o vetor de variáveis \mathbf{z} com i -ésima variável removida. Da proposição () temos que:

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

com $\mathbf{z} \in R(\mathbf{y}, \mathbf{z}) = \mathbf{z} = (z_1, z_2, \dots, z_n)^\top$; $z_i > 0$ se $y_i = 1$ ou $z_i \leq 0$ se $y_i = 0$. Utilizando o Lema C.0.3, temos que

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

sendo

$$K = \exp \left\{ \frac{1}{2} \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (z_k - \mu_k) \left(h_{kj} - \frac{h_{ki}h_{ij}}{1 - h_{ii}} \right) (z_j - \mu_j) \right\}$$

$$h_{ij} = \mathbf{x}_i^\top (\mathbf{v}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_j$$

$$\mu_i = \mathbf{x}_i^\top \mathbf{b} - \lambda\omega_i$$

Como K não depende de z_i , então a distribuição condicional completa de z_i pode ser escrita como

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

sendo $\text{Ind}(y_i, z_i) = \mathbf{I}(z_i > 0)\mathbf{I}(y_i = 1) + \mathbf{I}(z_i \leq 0)\mathbf{I}(y_i = 0)$. Podemos verificar que a fdp $\pi(z_i|\mathbf{z}_{-i}, \mathbf{y}, \boldsymbol{\omega}, \lambda)$ é proporcional ao núcleo de uma distribuição normal truncada. Logo, podemos escrever a distribuição condicional completa de cada elemento $z_i, i = 1, \dots, n$ da seguinte maneira:

$$z_i|\mathbf{z}_{-i}, y_i, \boldsymbol{\omega}, \lambda \propto \begin{cases} N(m_i, v_i)\mathbf{I}(z_i > 0) & \text{se } y_i = 1, \\ N(m_i, v_i)\mathbf{I}(z_i \leq 0) & \text{caso contrário,} \end{cases} \quad (\text{C.24})$$

sendo,

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

Podemos reescrever o parâmetro de localização m_i em função da média *a posteriori* de $\boldsymbol{\beta}$, $\mathbf{B} = \mathbf{V}[\mathbf{v}^{-1}\mathbf{b} + \mathbf{X}^\top(\mathbf{z} + \lambda\boldsymbol{\omega})]$, sendo $\mathbf{V} = (\mathbf{v}^{-1} + \mathbf{X}^\top \mathbf{X})^{-1}$. Deste modo, obteremos

uma representação que facilita na implementação do algoritmo de Gibbs. Escrevendo $h_{ij} = \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j$ temos

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

Utilizando o fato que $\sum_{j=1}^n \mathbf{x}_i^\top \mathbf{V} \mathbf{x}_j (z_j - \mathbf{x}_j^\top \mathbf{b} + \lambda \omega_j) = \mathbf{x}_i^\top \mathbf{B} - \mathbf{x}_i^\top \mathbf{b}$, temos que:

$$\begin{aligned} m_i &= \frac{1}{1 - h_i} [\mathbf{x}_i^\top \mathbf{B} - \mathbf{x}_i^\top \mathbf{b} - h_i (z_i - \mathbf{x}_i^\top \mathbf{b} - \mathbf{x}_i^\top \mathbf{b} + \lambda \omega_i)] + \mathbf{x}_i^\top \mathbf{b} - \lambda \omega_i \\ &= \mathbf{x}_i^\top \mathbf{B} - \lambda \omega_i - \frac{h_i}{1 - h_i} [z_i - (\mathbf{x}_i^\top \mathbf{B} - \lambda \omega_i)]. \end{aligned}$$

Visando facilitar o algoritmo de Gibbs, vamos optar por não calcular o vetor \mathbf{B} diretamente a cada atualização de algum z_i . Ou seja,

$$\begin{aligned} \mathbf{B} &= \mathbf{V} [\mathbf{v}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z} - \lambda \boldsymbol{\omega})] = \mathbf{V} \{ \mathbf{v}^{-1} \mathbf{b} + \mathbf{X}^\top [\mathbf{z}^{ant} + \mathbf{1}_i (z_i - z_i^{ant}) + \lambda \boldsymbol{\omega}] \} \\ \mathbf{B} &= \mathbf{V} [\mathbf{v}^{-1} \mathbf{b} + \mathbf{X}^\top (\mathbf{z}^{ant} - \lambda \boldsymbol{\omega})] + \mathbf{V} \mathbf{X}^\top \mathbf{1}_i (z_i - z_i^{ant}) \\ \mathbf{B} &= \mathbf{B}^{ant} + \mathbf{s}_i (z_i - z_i^{ant}), \end{aligned}$$

sendo \mathbf{B}^{ant} e z_i^{ant} denotam, respectivamente, os valores armazenados das atualizações anteriores de \mathbf{B} e z_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.

Para $\lambda = 0$ as distribuições condicionais completas dadas em C.24 se reduzem às distribuições condicionais completas de $(\mathbf{z}|\mathbf{y})$ para o modelo probito. ■

Proposição C.0.8. *Sob o modelo probito-assimétrico, temos que a distribuição $\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\omega})$ é dada por (3.26).*

Dado $\boldsymbol{\omega}$ o modelo probito-assimétrico se reduz ao modelo probito. Logo, para $\lambda = 0$ a prova da Proposição C.0.8 é obtida de maneira similar à prova da Proposição C.0.7.

Pseudo-Códigos

D.1 Distribuição normal-assimétrica

A distribuição e as propriedades da normal-assimétrica foi introduzida formalmente por Azzalini (1985). Esta distribuição representa a generalização da distribuição normal, na qual tem um parâmetro adicional que define a direção da assimetria da distribuição.

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

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

em que $\phi(\cdot)$ e $\Phi(\cdot)$ denotam, respectivamente, a função 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 $\Phi_{\mathcal{SN}}(\cdot; \mu, \sigma^2, \lambda)$.

Quando o vetor de parâmetros é $\theta = (0, 1 + \lambda^2, \lambda)$, 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 feitas por Sahu et al. (2003). Denotaremos essa função de distribuição acumulada (fda) por $\Phi_{CDS}(\cdot; \mu; \lambda)$, a função densidade é dada por:

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

Apresentaremos algumas propriedades da distribuição normal-assimétrica, D.1, que foram úteis para este trabalho.

Propriedades:

- a) Quando $\lambda = 0$, temos a distribuição $N(\mu, \sigma^2)$.
- b) Quando $\mu = 0$, $\sigma^2 = 1$ e $\lambda \rightarrow \infty$, temos a densidade *Half-Normal* (normal positiva), dada por:

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

Denotaremos por $\mathcal{HN}(1, 0)$.

- c) Sejam $\epsilon \sim N(0, 1)$ e $\omega \sim \mathcal{HN}(0, 1)$ variáveis aleatórias independentes, e $\lambda \in \mathbb{R}$, então:

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

- d) Se $Z|W = \omega \sim N(\mu + \lambda\omega, 1)$, $\omega \sim \mathcal{HN}(0, 1)$ e $\lambda \in \mathbb{R}$, então:

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

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

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

- f) 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.

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

D.2 Distribuição gaussiana-inversa

A distribuição gaussiana-inversa, denotada por GI. Pode-se encontrar mais detalhe no trabalho de Jorgensen (1982).

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

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

sendo $\mu > 0$ e $\lambda > 0$ os parâmetros de localização e de forma, respectivamente.

Em Devroye (1986) e Chhikara e Folks (1989) é representado o método de Michael, Schucany e Haas para amostrar de uma distribuição gaussiana-inversa usando transformação com múltiplas raízes.

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

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

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 por

1. Gera-se uma variável aleatória Y com distribuição qui-quadrado com 1 grau de liberdade;
2. Para cada valor gerado no passo 1, calcula-se X_1 utilizando a expressão D.3;
3. Se $U \leq \mu/(\mu + X_1)$, aceita-se X_1 com 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 estudado por Jorgensen (1982).

A 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 distribuição de probabilidade é dada por

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

sendo $\lambda \in \Re$, $\chi > 0$ e $\psi > 0$, os parâmetros da distribuição, e K a função modificada e Bessel de terceira espécie indexada por λ .

A função $K_\lambda(\omega)$ é dada por

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

Representação utilizadas por Devroye (1986) e Jorgensen (1982), respectivamente.

Os parâmetros da distribuição gaussiana-inversa generalizada têm as seguintes relações de dependência

- $\chi \geq 0$, $\psi > 0$ se $\lambda > 0$;
- $\chi > 0$, $\psi > 0$ e $\lambda = 0$;
- $\chi > 0$, $\psi \geq 0$, se $\lambda < 0$.

Condicionalada à dependência existente entre os parâmetros, podemos apresentar alguns casos especiais desta família de distribuições e algumas relações entre distribuições.

Alguns casos especiais da distribuições GIG

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

Alguns relações entre distribuições

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

D.4 Distribuição Kolmogorov-Smirnov

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

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

e a sua função densidade de probabilidade é dada por

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

Outra forma de escrever a função densidade de probabilidade da KS é dada por:

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

e a sua função de distribuição é dada por:

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

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* 19(6), 716–723.
- Anscombe, F. (1953). Contribution to the discussion of h. hotelling's paper. *Journal of the Royal Statistical Society* 15(2), 229–230.
- Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* 68(2), 357–363.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171–178.
- Azzalini, A. (2011). R package sn: The skew-normal and skew-t distributions (version 0.4-17). URL <http://azzalini.stat.unipd.it/SN> 20.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons.
- Bazán, J. L., M. D. Branco, H. Bolfarine, et al. (2006). A skew item response model. *Bayesian analysis* 1(4), 861–892.
- Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika* 52(3), 345–370.
- Chen, M.-H., D. K. Dey, e Q.-M. Shao (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* 94(448), 1172–1186.
- Chhikara, R. e J. Folks (1989). The inverse gaussian distribution. marcel decker. *Inc., New York*.
- Cho, H., J. G. Ibrahim, D. Sinha, e H. Zhu (2009). Bayesian case influence diagnostics for survival models. *Biometrics* 65(1), 116–124.

- Cordeiro, G. M. e Lima Neto, E. A. (2006). *Modelos Paramétricos*. Recife: Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática.
- Cox, D. R. e D. V. Hinkley (1986). *Theoretical statistics*. University Press, Cambridge.
- Cox, D. R. e E. J. Snell (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248–275.
- Devroye (1986). Non-uniform random variate generation. *NY: Springer*.
- Farias, R. B. A. (2007). Regressão binária bayesiana com o uso de variáveis auxiliares. Dissertação de Mestrado, Universidade de São Paulo.
- Geman, S. e D. Geman (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (6), 721–741.
- Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, Volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.
- Guerrero, V. M. e R. A. Johnson (1982). Use of the box-cox transformation with binary response models. *Biometrika* 69(2), 309–314.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57(1), 97–109.
- Holmes, C. C. e L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* 1(1), 145–168.
- Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) (2013). Disponível em: <http://portal.inep.gov.br>. Acesso em: Agosto de 2016.
- Jorgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*, Volume 9. Springer Science - Verlag.
- Littell, R. C., W. W. Stroup, G. A. Milliken, R. D. Wolfinger, e O. Schabenberger (2006). *SAS for mixed models*. SAS institute.
- Makridakis, S. e M. Hibon (1995). Evaluating accuracy(or error) measures.
- Nelder, J. e R. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A* 135, 370-384. *Journal of the Royal Statistical Society. Series A* 1972.
- Peng, F. e D. K. Dey (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics* 23(2), 199–213.
- Plummer, M., N. Best, K. Cowles, e K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R news* 6(1), 7–11.

- Rao, C. R., H. Foltys, e H. Toutenburg (1973). *Lineare statistische Methoden und ihre Anwendungen (Linear statistical inference and its applications, dt.-Übers.: Helge Toutenburg.) In dt. Sprache hrsg. v. Heinz Foltys.*
- Sahu, S. K., D. K. Dey, e M. D. Branco (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics* 31(2), 129–150.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics* 6(2), 461–464.
- Stefanski, L. A. (1991). A normal scale mixture representation of the logistic distribution. *Statistics & Probability Letters* 11(1), 69–70.
- Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association* 83(402), 426–431.
- Team, R. C. et al. (2013). R: A language and environment for statistical computing.
- UolEconomia (2016). Disponível em: <http://economia.uol.com.br>. Acesso em: Agosto de 2016.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society* 54(3), 426–482.
- Williams, D. (1984). Residuals in generalized linear models. Em *Proceedings of the 12th. International Biometrics Conference*, pp. 59–68.
- Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in statistics-Simulation and computation* 22(4), 1079–1106.