
Regressão binária usando ligações potência e
reversa de potência

Susan Alicia Chumbimune Anyosa

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

SUSAN ALICIA CHUMBIMUNE ANYOSA

REGRESSÃO BINÁRIA USANDO LIGAÇÕES POTÊNCIA E REVERSA DE POTÊNCIA

Dissertação apresentada ao Departamento de Estatística - Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Jorge Luis Bazán Guzmán

**São Carlos
Abril de 2017**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Susan Alicia Chumbimune Anyosa, realizada em 07/04/2017:

Prof. Dr. Jorge Luis Bazán Guzmán
USP

Prof. Dr. Caio Lucidius Naberezny Azevedo
UNICAMP

Prof. Dr. José Santos Romeo Núñez
Usach

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

SUSAN ALICIA CHUMBIMUNE ANYOSA

BINARY REGRESSION USING POWER AND REVERSAL POWER LINKS

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

Advisor: Prof. Dr. Jorge Luis Bazán Guzmán

São Carlos
April 2017

Aos estudantes que deixaram seus países para fazer pós-graduação em outros lares do mundo.

AGRADECIMENTOS

Expresso meu agradecimento a meus pais e irmãos, pelo apoio constante. Também aos amigos que fiz tanto em Lima como em São Carlos pelas lembranças que perdurarão na minha memória. Ao meu orientador prof. Jorge, pelos exemplos de dedicação e compromisso na pesquisa. Aos professores do programa PIPGEs pelos conhecimentos compartilhados. Aos funcionários da USP e da UFSCar, pelas orientações sempre cordiais ao longo de mais de dois anos.

Agradecimentos especiais a CAPES pelo apoio financeiro.

“And when a man is sad—I don’t mean because he has a toothache or has lost some money, but because he sees, for once in a way, how it all is with life and everything, and is sad in earnest—he always looks a little like an animal. He looks not only sad, but more right and more beautiful than usual.”
(Hermann Hesse)

RESUMO

O objetivo desta dissertação é estudar uma família de ligações assimétricas para modelos de regressão binária sob a abordagem bayesiana. Especificamente, apresenta-se a estimação dos parâmetros da família de modelos de regressão binária com funções de ligação potência e reversa de potência considerando o método de estimação Monte Carlo Hamiltoniano, na extensão No-U-Turn Sampler, e o método Metropolis-Hastings dentro de Gibbs. Além disso, estudam-se diferentes medidas de comparação de modelos, incluindo critérios de informação e de avaliação preditiva. Um estudo de simulação foi desenvolvido para estudar a acurácia e eficiência nos parâmetros estimados. Através da análise de dados educacionais, mostra-se que os modelos usando as ligações propostas apresentam melhor ajuste do que os modelos usando ligações tradicionais.

Palavras-chave: Regressão binária, ligação assimétrica, estimação bayesiana.

ABSTRACT

The aim of this dissertation is to study a family of asymmetric link functions for binary regression models under Bayesian approach. Specifically, we present the estimation of parameters of power and reversal power binary regression models considering Hamiltonian Monte Carlo method, on No-U-Turn Sampler extension, and Metropolis-Hastings within Gibbs sampling method. Furthermore, we study a wide variety of model comparison measures, including information criteria and measures of predictive evaluation. A simulation study was conducted in order to research accuracy and efficiency on estimated parameters. Through analysis of educational data we show that models using the proposed link functions perform better fit than models using standard links.

Keywords: Binary regression, asymmetric link, Bayesian estimation.

SUMÁRIO

1	INTRODUÇÃO	17
2	REGRESSÃO BINÁRIA COM FUNÇÕES DE LIGAÇÃO POTÊN- CIA E REVERSA DE POTÊNCIA	21
2.1	Regressão Binária	21
2.2	Propriedades de simetria e reversibilidade de distribuições de pro- babilidade	23
2.3	Distribuições potência e reversa de potência	25
2.4	Regressão binária com funções de ligação potência e reversa de potência	29
2.5	Probabilidade de sucesso para ligações potência e reversa de potência	36
2.6	Interpretação do parâmetro λ	39
3	ESTIMAÇÃO E AJUSTE DOS MODELOS	51
3.1	Estimação bayesiana	51
3.2	Comparação de modelos	57
4	ESTUDO DE SIMULAÇÃO	63
4.1	Considerações no estudo de simulação	63
4.2	Resultados	67
5	APLICAÇÃO	73
5.1	O conjunto de dados educacionais	73
5.2	Análise descritiva	74
5.3	Análise de regressão binária para o desempenho em matemática . .	76
6	CONSIDERAÇÕES FINAIS	85
	REFERÊNCIAS	87
ANEXO A	REVISÃO DO MÉTODO MONTE CARLO HAMILTONI- ANO	93
A.1	Dinâmica Hamiltoniana	93
A.2	MCMC na dinâmica Hamiltoniana	95

ANEXO B	CÓDIGOS ILUSTRATIVOS DOS MODELOS NO STAN E JAGS	97
B.1	Códigos no Stan	97
B.2	Códigos no JAGS	98
ANEXO C	RESULTADOS DA SIMULAÇÃO	101
ANEXO D	RESULTADOS DA APLICAÇÃO	111
D.1	Classificações segundo pontos de corte para a aplicação	111
D.2	Medidas de avaliação preditiva para a aplicação	112
D.3	Modelo de regressão binária com função de ligação potência logit	116

INTRODUÇÃO

A regressão logística tem sido aplicada largamente na análise de dados binários. Ela traz a vantagem da facilidade de interpretação dos parâmetros (JR; LEMESHOW; STURDIVANT, 2013). A função de ligação utilizada na regressão logística é o logit, a qual é uma função de ligação simétrica. Outras funções de ligação simétricas usadas comumente na regressão binária são o probit e o cauchit, que vêm das funções de distribuição acumuladas (fda) das distribuições Normal padrão e Cauchy padrão respectivamente. Maiores detalhes sobre a regressão binária podem ser encontrados em Collett (2003).

Nesse contexto, Czado e Santner (1992) pesquisaram os efeitos da má especificação na função de ligação. Eles chegaram à conclusão de que supor o logit como função de ligação no ajuste quando a função de ligação dos dados é outra, traz viés nas estimações dos parâmetros da regressão e nas probabilidades preditas. Além disso, mostraram que os efeitos são mais prejudiciais quando a função de ligação é afetada pela assimetria do que se fosse afetada pela curtose. As funções de ligação assimétricas mais usadas na prática e disponíveis nos pacotes estatísticos, são loglog e loglog complementar (ou cloglog), que vêm da distribuição Gumbel de Valor Extremo Máximo e da distribuição Gumbel de Valor Extremo Mínimo, respectivamente. Porém, essas funções de ligação não fornecem flexibilidade, pois mantêm uma assimetria de valor fixo. Chen, Dey e Shao (1999) também sinalizaram que uma função de ligação assimétrica é mais adequada quando existe muita diferença na quantidade de 1's e 0's na variável resposta.

Diante da necessidade de funções de ligação assimétricas que possam melhorar a modelagem em regressão binária, vários autores apresentaram novas propostas. Assim, por exemplo Aranda-Ordaz (1981), Guerrero e Johnson (1982) e Morgan (1983) consideraram famílias de um parâmetro; Prentice (1976) e Stukel (1988) introduziram famílias de dois parâmetros. O trabalho de Chen, Dey e Shao (1999) apresentou uma abordagem de variáveis latentes. Nagler (1994), Gupta e Gupta (2008), Kim, Chen e Dey (2008) e Wang e Dey (2010) introduziram um parâmetro de forma associado à assimetria.

Por outro lado, nos trabalhos de [Bazán, Romeo e Rodrigues \(2014\)](#), [Abanto-Valle, Bazán e Smith \(2014\)](#) e [Bazán et al. \(2016\)](#) foram apresentados os modelos de regressão binária com funções de ligação potência e reversa de potência. Eles se baseiam na ideia de introduzir um parâmetro adicional associado com as probabilidades de sucesso que as penaliza ou bonifica dependendo dos valores que pode assumir.

Neste trabalho, foi proposto estudar em detalhe as funções de ligação potência e reversa de potência já desenvolvidas em [Bazán, Romeo e Rodrigues \(2014\)](#), [Abanto-Valle, Bazán e Smith \(2014\)](#) e [Bazán et al. \(2016\)](#) para a regressão binária.

A estimação de modelos de regressão binária tradicionais (logit, probit e cloglog) desenvolvida sob abordagem clássica é apresentada em detalhe em [Agresti \(2007\)](#). Adicionalmente, a estimação bayesiana para estes modelos, pode ser revisada em [Dey, Gosh e Mallick \(2000\)](#) e [Ntzoufras \(2009\)](#).

Na estimação bayesiana, os métodos mais usados são baseados em Markov chain Monte Carlo (MCMC) ([NTZOUFRAS, 2009](#)). Eles fornecem algoritmos para gerar amostras correlacionadas dos parâmetros envolvidos que convergem em distribuição à distribuição a posteriori de interesse e foram originados pelo trabalho de [Metropolis et al. \(1953\)](#). Os algoritmos MCMC como random-walk Metropolis ([METROPOLIS et al., 1953](#)), Gibbs sampling ([GEMAN; GEMAN, 1984](#)) e Metropolis-Hastings dentro de Gibbs (MHG) são os mais utilizados para ajustar modelos de regressão binária na abordagem bayesiana.

[Duane et al. \(1987\)](#) propuseram o método Monte Carlo Hamiltoniano (MCH), também conhecido como Monte Carlo Híbrido, que foi aplicado na simulação molecular. O uso do MCH na estatística começou com o trabalho de [Neal \(1996\)](#) nas redes neurais bayesianas. No caso da regressão binária o MCH foi desenvolvido no trabalho de [Lan \(2013\)](#) para a regressão logística.

Nesta dissertação, para a estimação bayesiana, são considerados dois métodos: o No-U-Turn Sampler (NUTS), uma extensão do MCH, e o Metropolis-Hasting dentro de Gibbs. Diferentes métodos de comparação de modelos e medidas de avaliação preditiva também serão estudados.

O principal objetivo deste trabalho é apresentar uma abordagem bayesiana no que diz respeito à estimação de cinco modelos de regressão binária usando ligações potência e cinco modelos de regressão binária usando ligações reversa de potência, considerando dois métodos de estimação (MHG e NUTS). Adicionalmente são discutidos alguns pontos importantes como as propriedades desses modelos, em especial a interpretação do parâmetro λ , calculando sua influência na variação da probabilidade de sucesso, assim como sua relação com a assimetria das correspondentes distribuições estudadas. Além disso, realizou-se um estudo de simulação a fim de comparar os métodos de estimação estudados. Além do mais, aplicou-se a metodologia proposta num conjunto de dados reais, para exemplificar o contexto.

Este trabalho está organizado em seis capítulos e três anexos. No Capítulo 2, são apresentadas a regressão binária com funções de ligação simétricas e assimétricas, as propriedades de simetria e reversibilidade, de forma que definimos as distribuições potência e reversa de potência para depois apresentar suas respectivas funções de ligação. O Capítulo 3 é dedicado a detalhar a estimação e ajustes dos modelos propostos usando a aproximação bayesiana. Um estudo de simulação desenvolvido para estudar a acurácia dos parâmetros e a eficiência na estimação usando os algoritmos considerados, é apresentado no Capítulo 4. No Capítulo 5, uma aplicação para dados educacionais é desenvolvida modelando os dados considerando as ligações propostas. No Capítulo 6, são apresentadas as considerações finais deste trabalho assim como estudos sugeridos.

No Anexo A, apresenta-se uma revisão do método de simulação Monte Carlo Hamiltoniano. No Anexo B, os códigos de implementação dos modelos propostos são apresentados na linguagem dos softwares Stan e JAGS. No Anexo C, apresentam-se as medidas obtidas no estudo de simulação. No Anexo D, apresentam-se em detalhe mais resultados da aplicação para dados educacionais.

REGRESSÃO BINÁRIA COM FUNÇÕES DE LIGAÇÃO POTÊNCIA E REVERSA DE POTÊNCIA

Neste Capítulo os modelos de regressão binária com funções de ligação comuns (logit, probit, cauchit, cloglog e loglog) são apresentados. Além disso, as propriedades de simetria e reversibilidade presentes nas distribuições são estendidas, incluindo resultados e provas. Também as distribuições potência e reversa de potência são apresentadas, as propriedades dessas distribuições são detalhadas e as formas das distribuições de densidade e acumuladas são mostradas. Além do mais, é detalhada a regressão binária usando as funções de ligação potência e reversa de potência. Adicionalmente, é dada a interpretação do parâmetro de assimetria λ dos modelos propostos; assim como da relação que existe entre esse parâmetro e a probabilidade de sucesso na variável resposta.

2.1 Regressão Binária

Para os modelos de regressão binária considere $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$ um vetor $n \times 1$ de n variáveis aleatórias independentes binárias correspondentes as n observações da variável resposta. Seja $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{ik})'$ um vetor $(k+1) \times 1$ das covariáveis, $i = 1, \dots, n$ e seja \mathbf{X} a matriz $n \times (k+1)$ com linhas \mathbf{x}_i' . Seja também $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$ um vetor $(k+1) \times 1$ dos coeficientes de regressão associados às variáveis explicativas. Considere $Y_i = 1$ com probabilidade p_i e $Y_i = 0$ com probabilidade $1 - p_i$. Assim, o modelo de regressão binária é dada por:

$$Y_i \stackrel{ind.}{\sim} \text{Bernoulli}(p_i) \quad (2.1)$$

$$p_i = F(\eta_i) = F(\mathbf{x}_i' \boldsymbol{\beta}) \quad (2.2)$$

onde, $F(\cdot)$ denota a função de distribuição acumulada (fda), F^{-1} denomina-se função de ligação e $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$ é o i -ésimo preditor linear.

A função de ligação F^{-1} transforma o suporte de p no conjunto dos números reais \mathbb{R} . Quando o preditor linear η com valores em \mathbb{R} são avaliados na fda F (também conhecida como função de ligação inversa), os valores obtidos são garantidos como valores de probabilidade válidos que ficam entre 0 e 1. As funções de ligação consideradas comumente, ver por exemplo a função glm do software R (R Core Team, 2016), são apresentadas na Tabela 1.

Tabela 1 – Funções de ligação comuns.

Nome	Notação	F	F^{-1}
logit	L	$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$	$\log\left(\frac{p_i}{1-p_i}\right) = \eta_i$
probit	P	$p_i = \Phi(\eta_i)$	$\Phi^{-1}(p_i) = \eta_i$
cauchit	C	$p_i = 0,5 + \frac{\arctan(\eta_i)}{\pi}$	$\tan(\pi(p_i - 0,5)) = \eta_i$
cloglog	CLL	$p_i = 1 - \exp(-\exp(\eta_i))$	$\log(-\log(1-p_i)) = \eta_i$
loglog	LL	$p_i = \exp(-\exp(-\eta_i))$	$-\log(-\log(p_i)) = \eta_i$

A ligação logit é obtida da inversa da fda da distribuição Logística (L), a probit da Normal (N), a cauchit da Cauchy (C), a cloglog da Gumbel de Valor Mínimo (GVMi) e a loglog da Gumbel de Valor Máximo (GVMa), isto é mostrado em ambas as últimas colunas da Tabela 1.

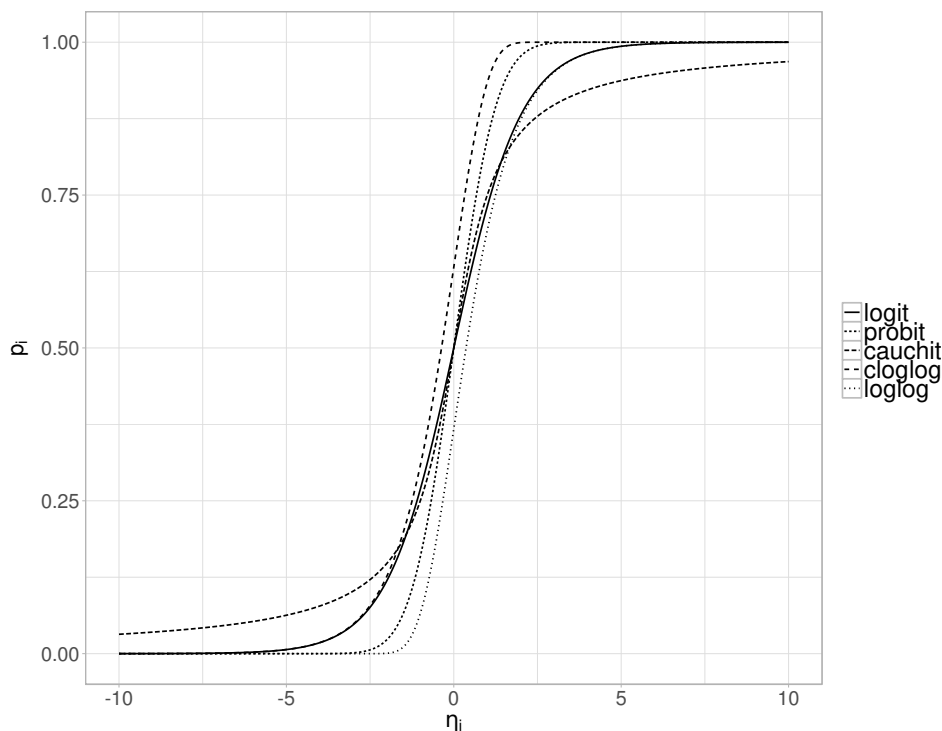


Figura 1 – Curva de resposta considerando funções de ligação comuns na regressão binária.

Na Figura 1, é mostrado o comportamento das diferentes ligações comuns para diversos valores do preditor linear. Note que as ligações logit, probit e cauchit são simétricas em torno de $p = 0,5$, para o valor $\eta = 0$. De outra forma, as ligações cloglog e loglog são assimétricas em torno de $p = 0,5$, pois nesse valor $\eta = -0,3665$ na ligação cloglog e $\eta = 0,3665$ na ligação loglog. Observe-se que cloglog e logit são difíceis de distinguir entre si quando η tem valores pequenos, assim também loglog e logit são difíceis de distinguir quando η tem valores grandes.

2.2 Propriedades de simetria e reversibilidade de distribuições de probabilidade

Bazán *et al.* (2016) definem as propriedades necessárias a serem consideradas quando são usadas diferentes funções de ligação.

Propriedade de reversibilidade. Se $S \sim F(\cdot)$ (ou $S \sim f(\cdot)$), diz-se que a distribuição de S satisfaz a propriedade de reversibilidade se $-S \sim G(\cdot) \equiv 1 - F(-\cdot)$. Nesse caso $G(\cdot)$ é nomeada a distribuição reversa de $F(\cdot)$.

Propriedade de simetria. Seja $S \sim F(\cdot)$ (ou $S \sim f(\cdot)$), diz-se que a distribuição de S é simétrica se existe um valor x_0 de forma que $f(x_0 - \gamma) = f(x_0 + \gamma)$ para qualquer número real γ . Assim, neste caso S e $-S$ tem a mesma distribuição, e então S não satisfaz a propriedade de reversibilidade.

A mediana e a média (se existir) de uma distribuição simétrica ocorrem no ponto de simetria x_0 . Daqui em diante, ao falar de simetria (ou assimetria), será referindo-se à simetria (ou assimetria) em torno ao ponto $x_0 = 0$.

A seguir serão usadas essas propriedades para avaliar as ligações comuns na regressão binária.

Resultado 1. As distribuições Logística, Normal e Cauchy satisfazem a *propriedade de simetria*.

Prova. Para distribuições simétricas têm-se o seguinte:

Sem perda de generalidade considere $S \sim \text{Logística}(0,1)$:

$$f(s) = \frac{\exp(-s)}{(1 + \exp(-s))^2}.$$

Na distribuição Logística(0,1), $x_0 = 0$ é a mediana. Consequentemente:

$$f(x_0 - \gamma) = f(-\gamma) = \frac{\exp(\gamma)}{(1 + \exp(\gamma))^2},$$

$$f(x_0 + \gamma) = f(\gamma) = \frac{\exp(-\gamma)}{(1 + \exp(-\gamma))^2} = \frac{\frac{1}{\exp(\gamma)}}{1 + \frac{2}{\exp(\gamma)} + \frac{1}{\exp(2\gamma)}} = \frac{\exp(\gamma)}{1 + 2\exp(\gamma) + \exp(2\gamma)} = \frac{\exp(\gamma)}{(1 + \exp(\gamma))^2}.$$

Então $f(x_0 - \gamma) = f(x_0 + \gamma)$, portanto, S é simétrica.

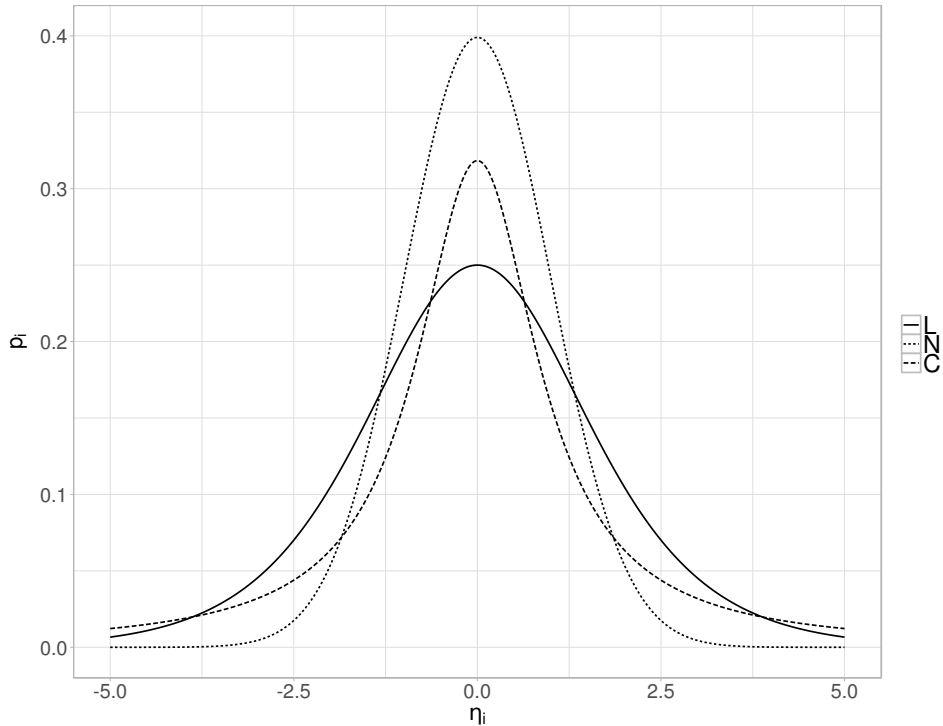


Figura 2 – Fdp das distribuições simétricas Logística, Normal e Cauchy.

Na Figura 2 das fdp é visto claramente que $f(s) = f(-s)$, para as distribuições citadas.

Resultado 2. As distribuições Gumbel de Valor Mínimo e Gumbel de Valor Máximo, F^* e F^{**} respectivamente, não satisfazem a *propriedade de simetria*. Porém satisfazem a *propriedade de reversibilidade* uma vez que $F^{**}(s) + F^*(-s) = 1$ e $f^{**}(s) = f^*(-s)$.

Prova. Se S tem distribuição Gumbel de Valor Máximo então:

$$F^{**}(s) = e^{-e^{-s}}.$$

Enquanto que, se S tem distribuição Gumbel de Valor Mínimo, então:

$$F^*(s) = 1 - e^{-e^s}.$$

Note-se neste caso que:

$$F^{**}(s) + F^*(-s) = e^{-e^{-s}} + 1 - e^{-e^{-s}} = 1.$$

Consequentemente, uma é a reversa da outra. É trivial mostrar que:

$$f^*(x_0 - \gamma) \neq f^*(x_0 + \gamma),$$

e:

$$f^{**}(x_0 - \gamma) \neq f^{**}(x_0 + \gamma),$$

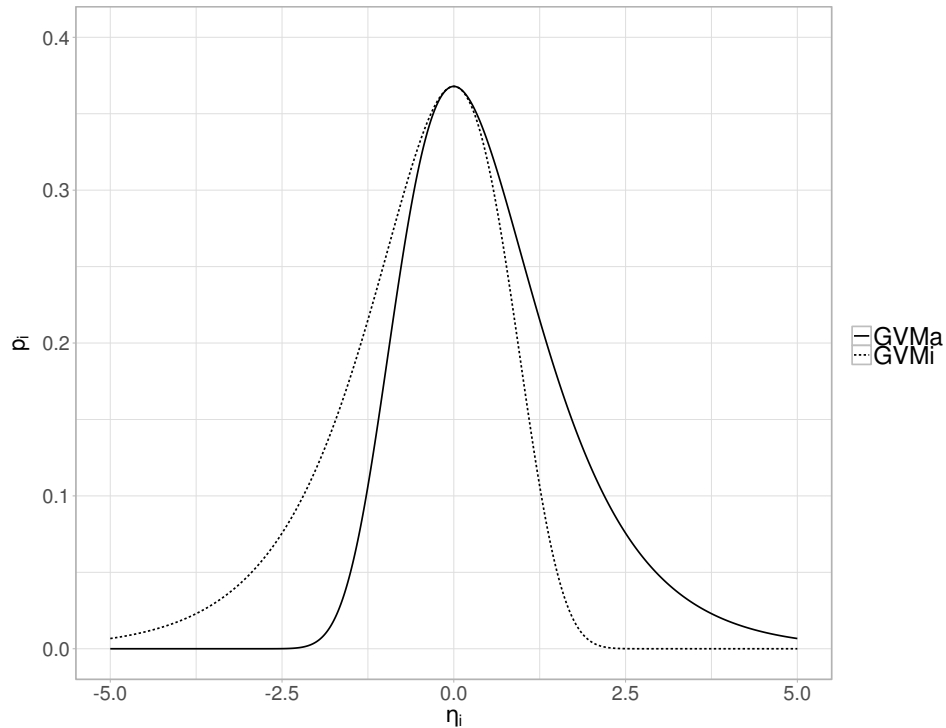


Figura 3 – Fdp das distribuições assimétricas Gumbel de Valor Máximo e Gumbel de Valor Mínimo.

Na Figura 3 é visto que nas fdp das distribuições Gumbel de Valor Máximo e Gumbel de Valor Mínimo, $f^{**}(s) = f^{*}(-s)$.

Comentário. Verificar a propriedade de simetria numa distribuição é importante. Caso ela seja verificada, a distribuição é simétrica, e, conseqüentemente, não apresenta reversibilidade. Se não for verificada a simetria, é possível propor uma distribuição reversa a ela.

2.3 Distribuições potência e reversa de potência

Como apresentado em [Abanto-Valle, Bazán e Smith \(2014\)](#) e [Bazán et al. \(2016\)](#), a variável aleatória (va) univariada S segue a distribuição potência $S \sim P(\mu, \sigma^2, \lambda)$, onde $-\infty < \mu < \infty$ é o parâmetro de localização, $\sigma^2 > 0$ é o parâmetro de escala e $\lambda > 0$ é o parâmetro de forma. A distribuição S tem a seguinte função de densidade de probabilidade (fdp):

$$f_P(s | \mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} g\left(\frac{s-\mu}{\sigma}\right) \left[G\left(\frac{s-\mu}{\sigma}\right)\right]^{\lambda-1}, \quad (2.3)$$

onde $g(\cdot)$ é uma fdp unimodal, log concava de suporte real e $G(\cdot)$ é a fda de $g(\cdot)$, nomeada de distribuição base que pode ser qualquer fda simétrica ou assimétrica. Observamos que se $\lambda = 1$, a densidade de S é reduzida à densidade de $g(\mu, \sigma^2)$.

Considera-se também o caso da distribuição potência padrão, denotada como $Z \sim P(\lambda)$, que é obtida quando $\mu = 0$ e $\sigma = 1$. A fdp de Z é dada por:

$$f_P(z | \lambda) = \lambda g(z) [G(z)]^{\lambda-1}, \quad (2.4)$$

e a fda é dada por:

$$F_P(z) = [G(z)]^\lambda. \quad (2.5)$$

A fdp da distribuição reversa de potência é dada por:

$$f_{RP}(z | \lambda) = \lambda g(z) [G(-z)]^{\lambda-1}, \quad (2.6)$$

e a fda é dada por:

$$F_{RP}(z) = 1 - [G(-z)]^\lambda. \quad (2.7)$$

A versão localização-escala da reversa de potência $S \sim RP(\mu, \sigma^2, \lambda)$ é dada por:

$$f_{RP}(s | \mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} g\left(\frac{s-\mu}{\sigma}\right) \left[G\left(-\left(\frac{s-\mu}{\sigma}\right)\right) \right]^{\lambda-1} \quad (2.8)$$

Considerando o anterior, tem-se o seguinte resultado:

Resultado 3. É visto que $Z \sim F_P(\cdot)$ satisfaz a propriedade de reversibilidade com $F_{RP}(z) = 1 - G(-z)^\lambda$ sendo sua distribuição reversa. Ao mesmo tempo considere Z uma va com fda $F_{RP}(z) = 1 - G(-z)^\lambda$ nomeada reversa de potência padrão, denotada como $Z \sim RP(\lambda)$, onde G é uma distribuição base e λ é o parâmetro de forma; $F_{RP}(\cdot)$ satisfaz a propriedade de reversibilidade com $F_{RP}(z) = G(z)^\lambda$ sendo sua distribuição reversa.

Prova. Lembrar que a propriedade de reversibilidade indica que $F^{**}(z) + F^*(-z) = 1$, onde F^{**} é a distribuição de interesse e F^* é sua distribuição reversa. Isto é cumprido nos casos das distribuições potência e reversa de potência.

Por exemplo, seja $F^{**} = \left[\frac{e^z}{1+e^z} \right]^\lambda$ e $F^* = 1 - \left[\frac{e^{-z}}{1+e^{-z}} \right]^\lambda$, as funções de distribuição acumulada das distribuições Potência Logística e Reversa de Potência Logística respectivamente. Então verifica-se que $F^{**}(z) + F^*(-z) = \left[\frac{e^z}{1+e^z} \right]^\lambda + 1 - \left[\frac{e^z}{1+e^z} \right]^\lambda = 1$.

Observações. É fácil verificar que:

- F_P e F_{RP} não são ponto-simétricas uma vez que $F_P(-z) \neq 1 - F_P(z)$ ou $F_{RP}(-z) \neq 1 - F_{RP}(z)$ para $\lambda \neq 1$.

- $F_P(\pm z) + F_{RP}(\mp z) = 1$ o que quer dizer que as duas distribuições são diferentes mas estão relacionadas porque uma é reversa da outra.
- Se $\lambda = 1$ então $F_P(z) = G(z) = F_{RP}(z)$. Isto quer dizer que $G(\cdot)$ é um caso particular das duas distribuições ou distribuição base.
- Se $Z \sim F_P(\cdot)$, então $-Z \sim F_{RP}(\cdot)$.

Resultado 4. É visto que se $Z \sim F_P(\lambda)$ ou $Z \sim F_{RP}(\lambda)$, então Z não é simétrica se $\lambda \neq 1$.

Prova. Lembrar que na equação (2.4) tem-se que a fdp da distribuição potência padrão é $f_P(z | \lambda) = \lambda g(z) [G(z)]^{\lambda-1}$, e então avaliando a fdp em $-z$: $f_P(-z | \lambda) = \lambda g(-z) [G(-z)]^{\lambda-1}$. Consequentemente:

$$f_P(z | \lambda) \neq f_P(-z | \lambda).$$

Embora $g(z) = g(-z)$ nas distribuições simétricas, tem-se que $G(z) \neq G(-z)$. Pode-se concluir que ambas as distribuições potência e reversa de potência não são simétricas em torno de zero.

Neste trabalho consideram-se versões de potência e reversa de potência para as cinco distribuições comuns (Logística, Normal, Cauchy, Gumbel de Valor Mínimo e Gumbel de Valor Máximo) que dão origem as funções de ligação comuns. Assim também, consideram-se cinco distribuições potência e cinco distribuições reversa de potência. Isto é apresentado nas Tabelas 2 e 3.

Também sabe-se que as distribuições base das funções de ligação logit, probit, cauchit, cloglog e loglog são casos particulares das versões potência e reversa de potência porque nelas $\lambda = 1$. Assim é possível considerar um conjunto amplo de distribuições com as quais pode-se propor diferentes ligações potência e reversa de potência, usando as correspondentes fda.

Tabela 2 – Funções densidade de probabilidade e funções de distribuição acumulada das distribuições potência.

Distribuição ¹	Densidade	Acumulada
PL	$f_{PL}(z) = \lambda \left[\frac{e^{-z}}{(1+e^{-z})^2} \right] \left[\frac{e^z}{1+e^z} \right]^{\lambda-1}$	$F_{PL}(z) = \left[\frac{e^z}{1+e^z} \right]^\lambda$
PN	$f_{PN}(z) = \lambda \left[\frac{e^{-z^2}}{\sqrt{2\pi}} \right] [\Phi(z)]^{\lambda-1}$	$F_{PN}(z) = [\Phi(z)]^\lambda$
PC	$f_{PC}(z) = \lambda \left[\frac{1}{\pi(1+z^2)} \right] \left[\frac{1}{\pi} \arctan(z) + \frac{1}{2} \right]^{\lambda-1}$	$F_{PC}(z) = \left[\frac{1}{\pi} \arctan(z) + \frac{1}{2} \right]^\lambda$
PGVMi	$f_{PGVMi}(z) = \lambda \left[e^{-(z+e^z)} \right] [1 - e^{-e^z}]^{\lambda-1}$	$F_{PGVMi}(z) = [1 - e^{-e^z}]^\lambda$
PGVMa	$f_{PGVMa}(z) = \lambda \left[e^{-(z+e^{-z})} \right] [e^{-e^{-z}}]^{\lambda-1}$	$F_{PGVMa}(z) = [e^{-e^{-z}}]^\lambda$

¹ PL: Potência Logística, PN: Potência Normal, PC: Potência Cauchy, PGVMi: Potência Gumbel de Valor Mínimo, PGVMa: Potência Gumbel de Valor Máximo.

Tabela 3 – Funções densidade de probabilidade e funções de distribuição acumulada das distribuições reversa de potência.

Distribuição ¹	Densidade	Acumulada
RPL	$f_{RPL}(z) = \lambda \left[\frac{e^{-z}}{(1+e^{-z})^2} \right] \left[\frac{e^{-z}}{1+e^{-z}} \right]^{\lambda-1}$	$F_{RPL}(z) = 1 - \left[\frac{e^{-z}}{1+e^{-z}} \right]^{\lambda}$
RPN	$f_{RPN}(z) = \lambda \left[\frac{e^{-z^2}}{\sqrt{2\pi}} \right] [\Phi(-z)]^{\lambda-1}$	$F_{RPN}(z) = 1 - [\Phi(-z)]^{\lambda}$
RPC	$f_{RPC}(z) = \lambda \left[\frac{1}{\pi(1+z^2)} \right] \left[\frac{1}{\pi} \arctan(-z) + \frac{1}{2} \right]^{\lambda-1}$	$F_{RPC}(z) = 1 - \left[\frac{1}{\pi} \arctan(-z) + \frac{1}{2} \right]^{\lambda}$
RPGVMi	$f_{RPGVMi}(z) = \lambda \left[e^{-(z+e^z)} \right] \left[1 - e^{-e^{-z}} \right]^{\lambda-1}$	$F_{RPGVMi}(z) = 1 - \left[1 - e^{-e^{-z}} \right]^{\lambda}$
RPGVMa	$f_{RPGVMa}(z) = \lambda \left[e^{-(z+e^{-z})} \right] \left[e^{-e^z} \right]^{\lambda-1}$	$F_{RPGVMa}(z) = 1 - \left[e^{-e^z} \right]^{\lambda}$

¹ PRL: Reversa de Potência Logística, RPN: Reversa de Potência Normal, RPC: Reversa de Potência Cauchy, RPGVMi: Reversa de Potência Gumbel de Valor Mínimo, RPGVMa: Reversa de Potência Gumbel de Valor Máximo.

A distribuição Potência Logística é conhecida também como a distribuição logística generalizada Tipo-I (JOHNSON; KOTZ; BALAKRISHNAN, 1995). A função de ligação obtida da distribuição Reversa de Potência Logística foi nomeada como scobit por Nagler (1994). A distribuição Potência Normal foi apresentada por Gupta e Gupta (2008). A distribuição Reversa de Potência Normal foi apresentada por Bazán, Romeo e Rodrigues (2014). Essas duas são casos particulares da distribuição beta-normal como sinalizado em Bazán *et al.* (2016).

As distribuições Potência Gumbel de Valor Mínimo, Reversa de Potência Gumbel de Valor Mínimo, Potência Gumbel de Valor Máximo e Reversa de Potência Gumbel de Valor Máximo foram propostas em Abanto-Valle, Bazán e Smith (2014). As distribuições Potência Cauchy e Reversa de Potência Cauchy foram apresentadas em Bazán *et al.* (2016).

Considerando a propriedade de reversibilidade é possível propor outras distribuições potência e reversa de potência.

2.4 Regressão binária com funções de ligação potência e reversa de potência

Para os modelos de regressão binária com função de ligação potência e reversa de potência considere-se $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$ um vetor $n \times 1$ de n variáveis aleatórias independentes binárias correspondentes as n observações da variável resposta. Seja $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{ik})'$ um vetor $(k+1) \times 1$ das covariáveis, $i = 1, \dots, n$ e seja \mathbf{X} a matriz $n \times (k+1)$ com linhas \mathbf{x}_i' . Seja também $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$ um vetor $(k+1) \times 1$ dos coeficientes de regressão associados às variáveis explicativas. Considere-se $Y_i = 1$ com probabilidade p_i e $Y_i = 0$ com probabilidade $1 - p_i$. Assim, o modelo de regressão binária é dado por:

$$Y_i \stackrel{ind.}{\sim} Bernoulli(p_i) \quad (2.9)$$

$$p_i = F_\lambda(\eta_i) = F_\lambda(\mathbf{x}_i' \boldsymbol{\beta}) \quad (2.10)$$

onde, $F_\lambda(\cdot)$ denota a fda de alguma distribuição potência ou reversa de potência, F_λ^{-1} denomina-se função de ligação potência ou reversa de potência e $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$ é o i -ésimo preditor linear.

Na Tabela 4 são apresentadas as cinco funções de ligação potência e as cinco funções de ligação reversa de potência estudadas neste trabalho.

Tabela 4 – Funções de ligação potência e reversa de potência.

Nome	Notação	F_λ	$F_{\lambda=1}^{-1}$
Potência logit	PL	$p_i = \left(\frac{e^{\eta_i}}{1+e^{\eta_i}} \right)^\lambda$	$\log \left(\frac{p_i}{1-p_i} \right) = \eta_i$
Reversa de potência logit	RPL	$p_i = 1 - \left(\frac{e^{-\eta_i}}{1+e^{-\eta_i}} \right)^\lambda$	
Potência probit	PP	$p_i = (\Phi(\eta_i))^\lambda$	$\Phi^{-1}(p_i) = \eta_i$
Reversa de potência probit	RPP	$p_i = 1 - (\Phi(-\eta_i))^\lambda$	
Potência cauchit	PC	$p_i = \left(0,5 + \frac{\arctan(\eta_i)}{\pi} \right)^\lambda$	$\tan(\pi(p_i - 0,5)) = \eta_i$
Reversa de potência cauchit	RPC	$p_i = 1 - \left(0,5 + \frac{\arctan(-\eta_i)}{\pi} \right)^\lambda$	
Potência cloglog	PCLL	$p_i = \left(1 - e^{-e^{\eta_i}} \right)^\lambda$	$\log(-\log(1-p_i)) = \eta_i$
Reversa de potência cloglog	RPCLL	$p_i = 1 - \left(1 - e^{-e^{-\eta_i}} \right)^\lambda$	
Potência loglog	PLL	$p_i = \left(e^{-e^{-\eta_i}} \right)^\lambda$	$-\log(-\log(p_i)) = \eta_i$
Reversa de potência loglog	RPLL	$p_i = 1 - \left(e^{-e^{\eta_i}} \right)^\lambda$	

Para ter uma ideia das diferenças entre as funções de ligação, nas Figuras 4 a 13 são apresentadas as curvas das funções de resposta considerando as diferentes ligações potência e reversa de potência para diferentes valores de $\lambda = 1/5; 1/3; 1; 2; 3$.

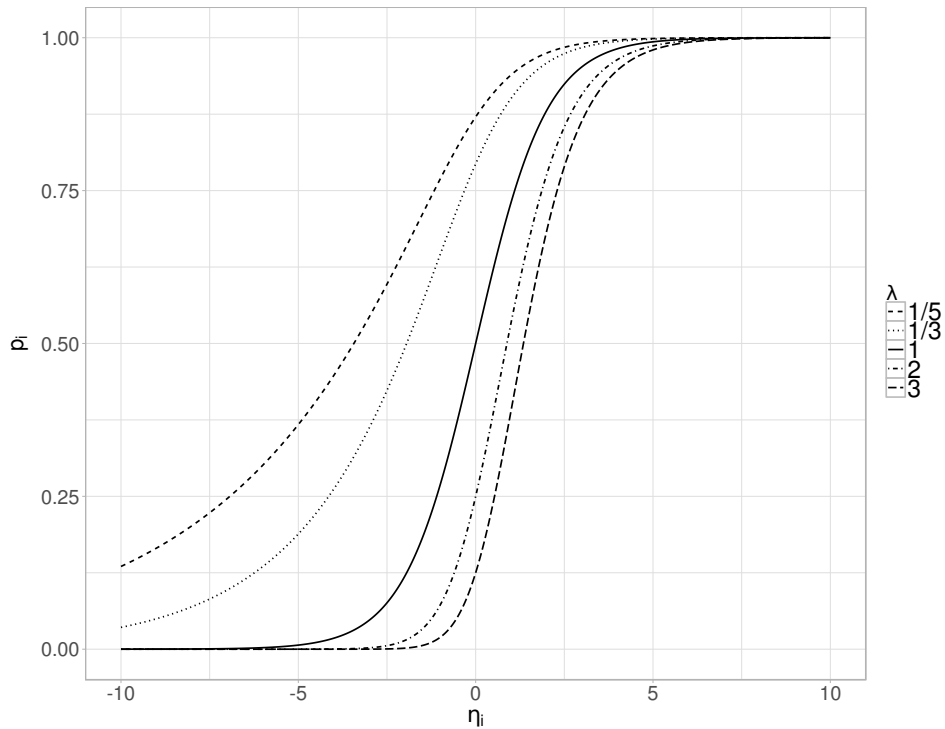


Figura 4 – Curva de resposta considerando ligação PL.

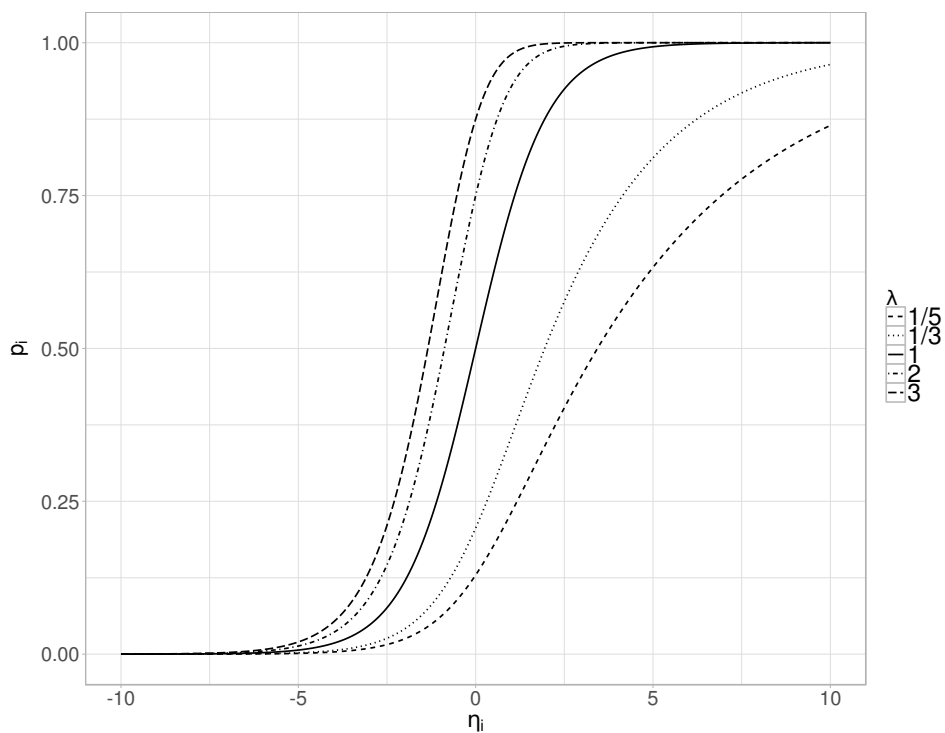


Figura 5 – Curva de resposta considerando ligação RPL.

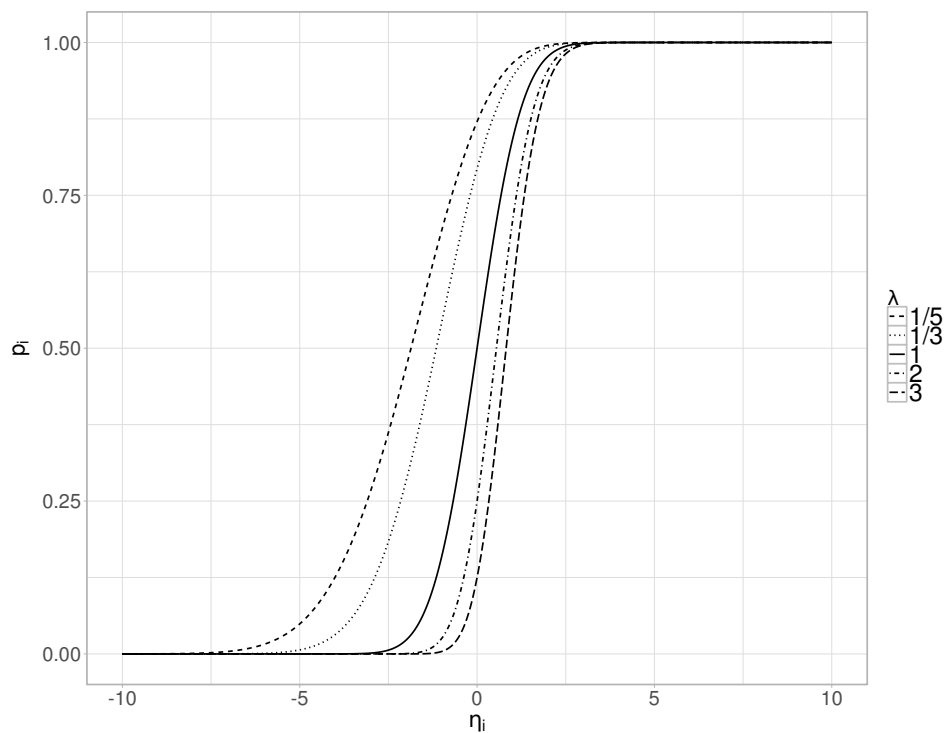


Figura 6 – Curva de resposta considerando ligação PP.

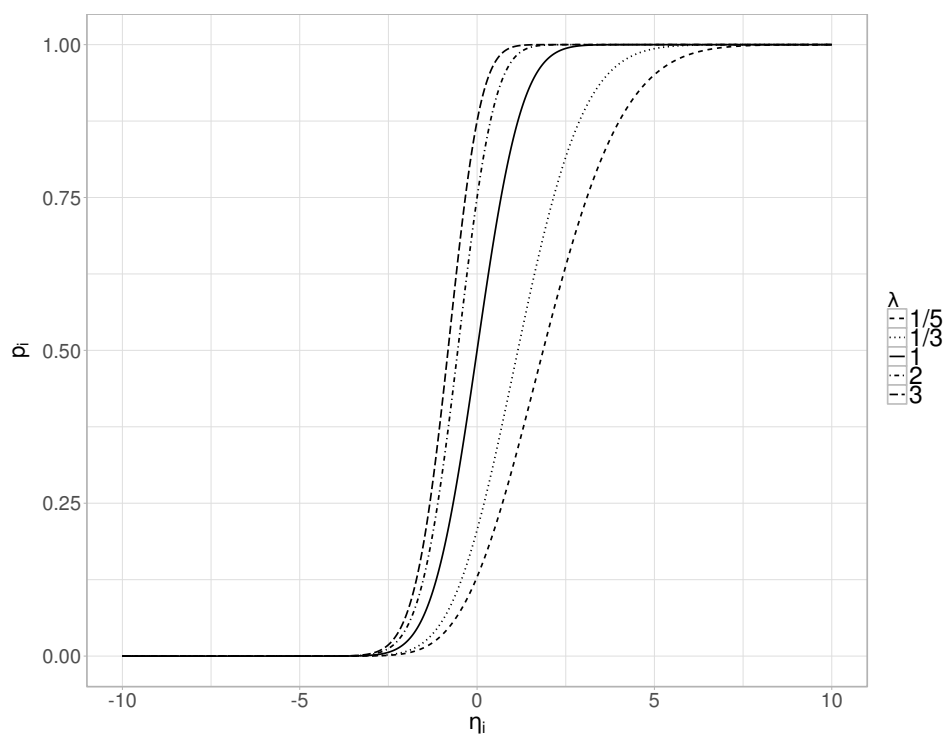


Figura 7 – Curva de resposta considerando ligação RPP.

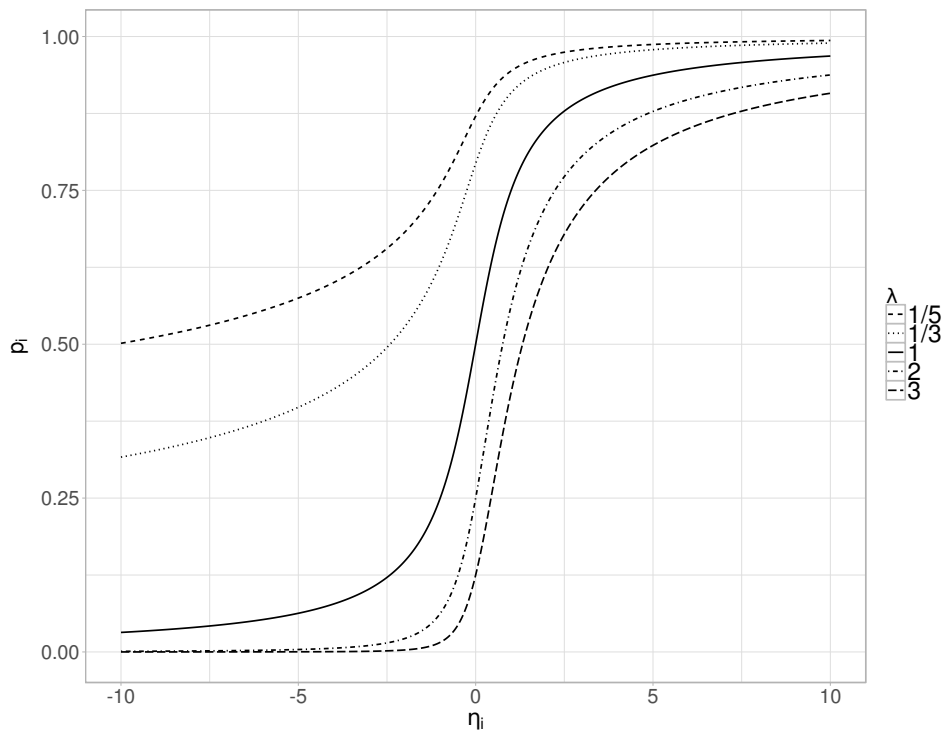


Figura 8 – Curva de resposta considerando ligação PC.

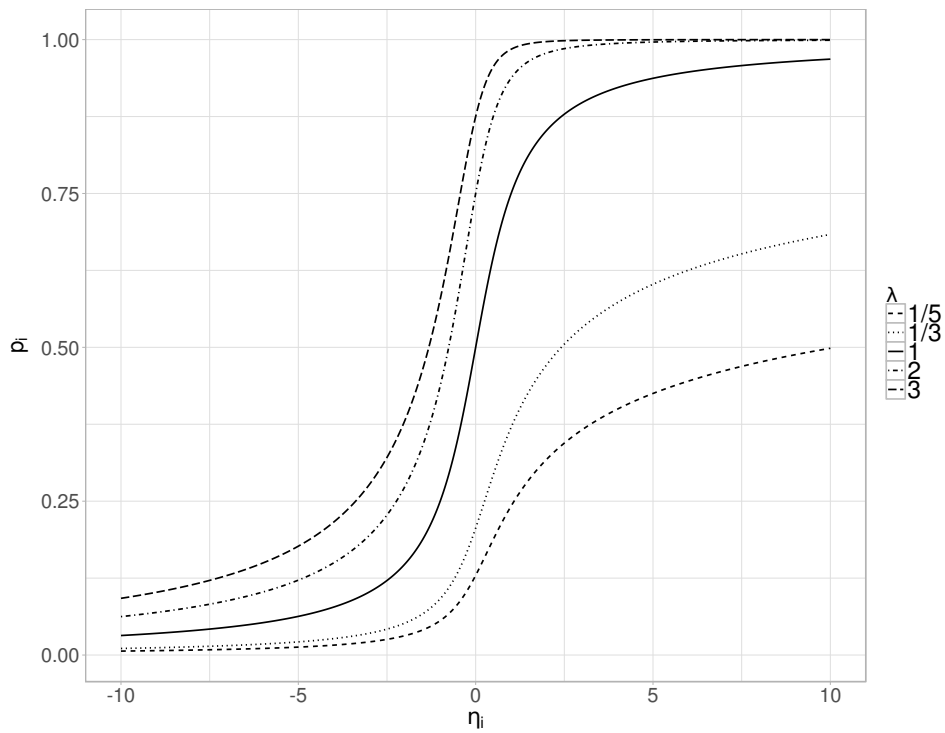


Figura 9 – Curva de resposta considerando ligação RPC.

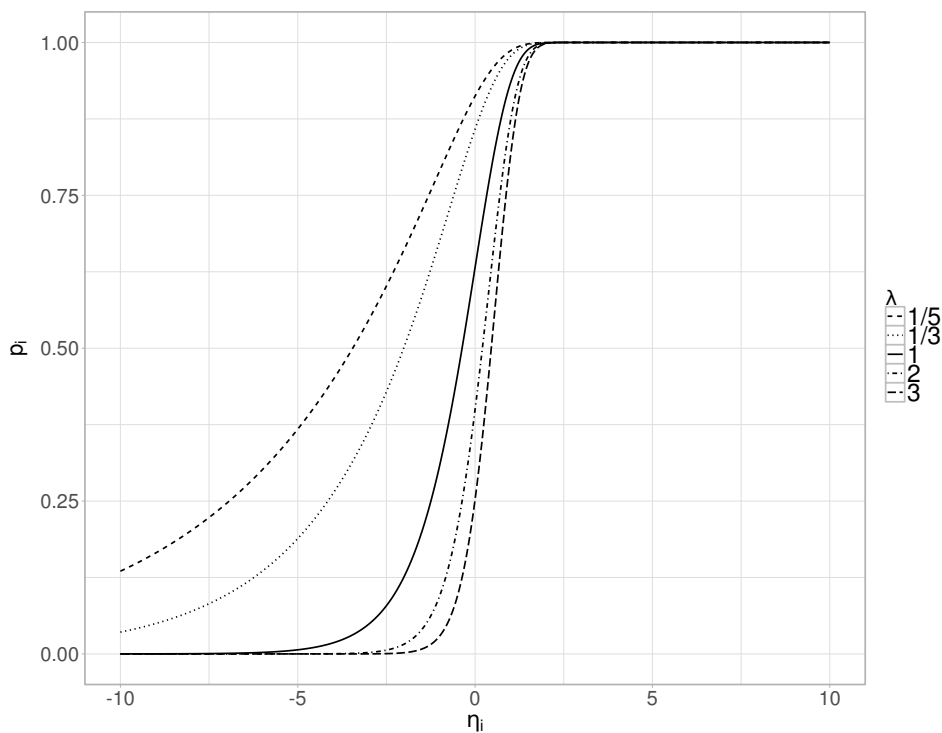


Figura 10 – Curva de resposta considerando ligação PCLL.

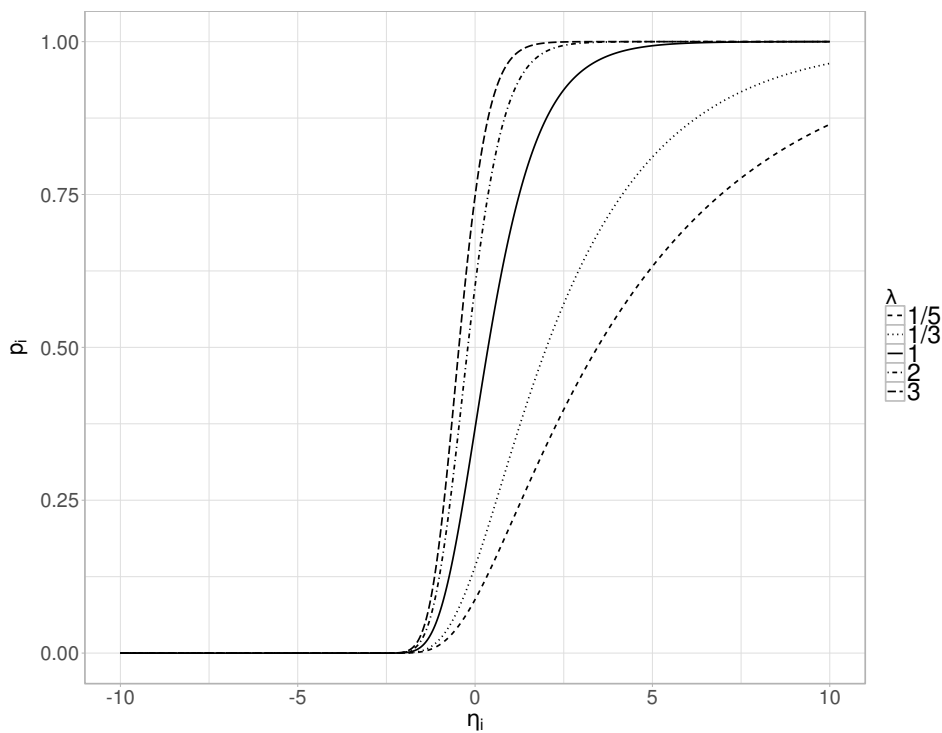


Figura 11 – Curva de resposta considerando ligação RPCLL.

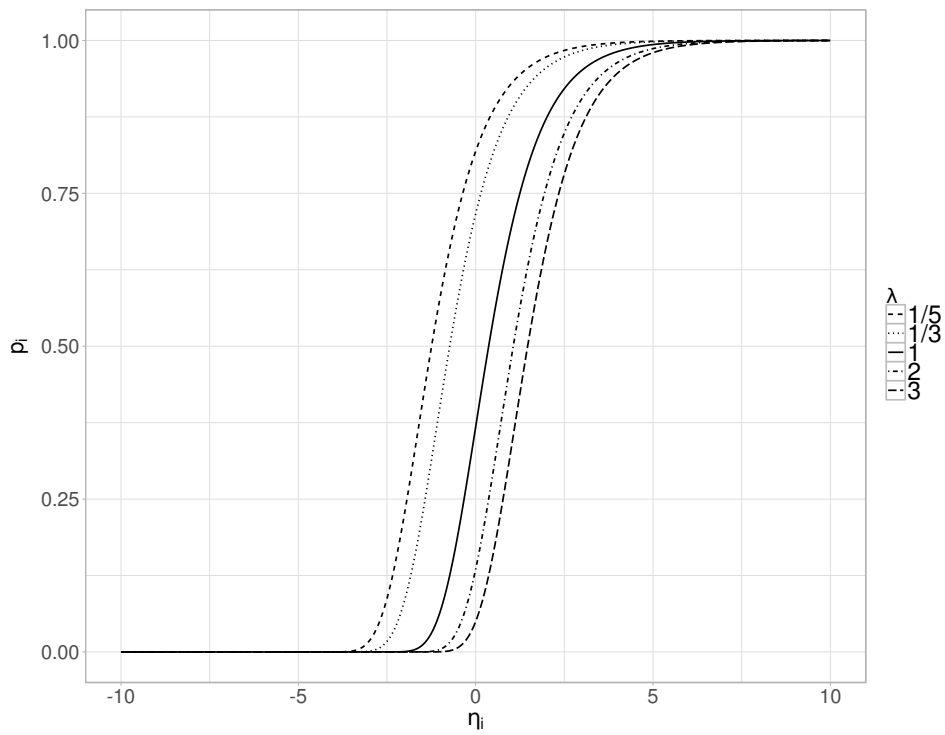


Figura 12 – Curva de resposta considerando ligação PLL.

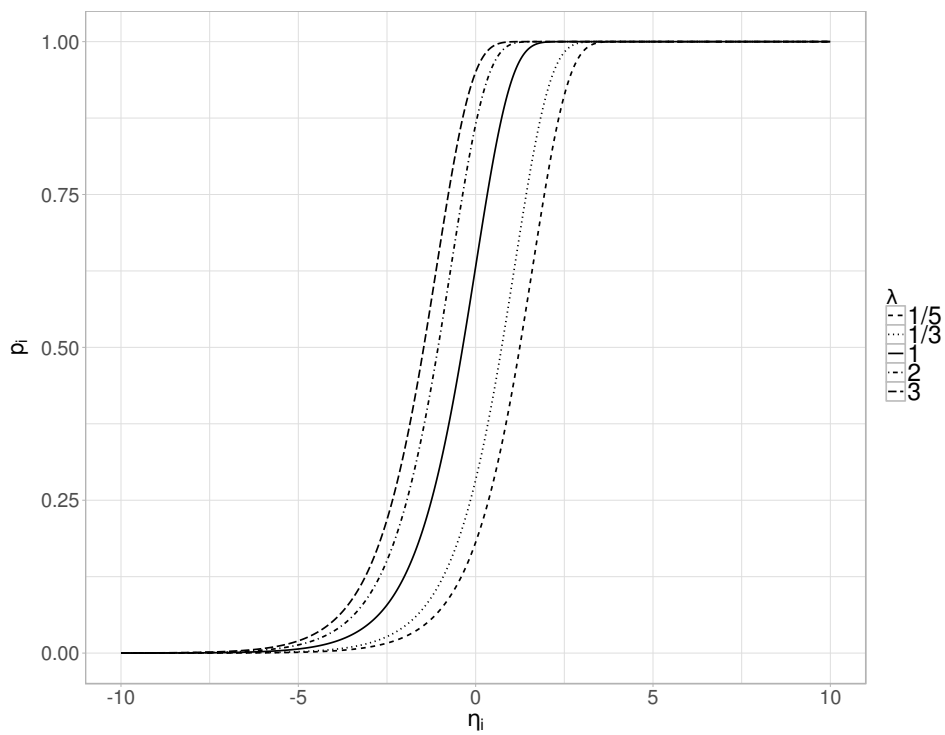


Figura 13 – Curva de resposta considerando ligação RPLL.

Em todas as figuras a linha em preto ($\lambda = 1$), representa a curva de resposta da distribuição de base das ligações comuns apresentadas na Figura 1. Em todos os casos note que para $\lambda \neq 1$ todas as funções de ligação são assimétricas. Adicionalmente pode ser observado que:

- As dez funções de ligação propostas são diferentes umas das outras, e dependem do valor de λ . Em geral, note que nas ligações potência, a curva de resposta para $\lambda < 1$ está a esquerda da curva de resposta base; e a curva de resposta para $\lambda > 1$ está a direita da curva de resposta base. Por sua vez, quando observam-se as ligações reversa de potência, o efeito é o contrário.
- Note também que as curvas de resposta mais flexíveis correspondem as ligações usando a distribuição de base Cauchy, a menos flexíveis correspondem às ligações usando a distribuição de base Gumbel de Valor Máximo.
- Além disso, observe que existe reversibilidade entre as funções de ligação potência e sua correspondente função de ligação reversa, no sentido que uma é o espelho da outra.
- Para estudar mais as diferenças entre as ligações propostas, na seção 2.6 estuda-se o efeito do parâmetro λ na mudança das curvas de probabilidade, assim como a assimetria que cada uma das distribuições apresenta.

Na seguinte seção é analisada a probabilidade de sucesso e sua relação com o parâmetro λ .

2.5 Probabilidade de sucesso para ligações potência e reversa de potência

Como foi indicado na introdução, [Chen, Dey e Shao \(1999\)](#) sinalizaram que uma função de ligação assimétrica é mais adequada quando existe muita diferença na quantidade de 1's e 0's na variável resposta. Para verificar isso foi realizado um breve estudo de simulação. Dessa forma, foram simuladas 100 réplicas de conjuntos de dados de 1000 observações das distribuições dos modelos de regressão binária com funções de ligação potência logit (PL) e reversa de potência logit (RPL). Os valores de β foram fixados a $\beta_0 = 0$ e $\beta_1 = 1$. Ao só considerar uma covariável na matriz \mathbf{X} , so foi gerada x_1 . Os valores da covariável foram gerados aleatoriamente da distribuição $Uniforme(-3,3)$. Foram considerados diferentes cenários dos valores de λ com $\lambda = 1/5; 1/3; 1; 2; 3$. Para cada caso obteve-se a proporção de 1's da variável resposta gerada. São apresentados os gráficos histograma e boxplot das proporções de 1's para cada cenário ao longo das 100 réplicas

Nas Figuras 14 e 15 pode-se observar que para $0 < \lambda < 1$ o histograma e o boxplot apresentam valores maiores das proporções de 1's. Como esperado, quando $\lambda = 1$ é observado o histograma e o boxplot com valores das proporções de 1's próximos a 0,5. No caso dos valores de $\lambda > 1$ o histograma e o boxplot apresentam valores menores das proporções de 1's.

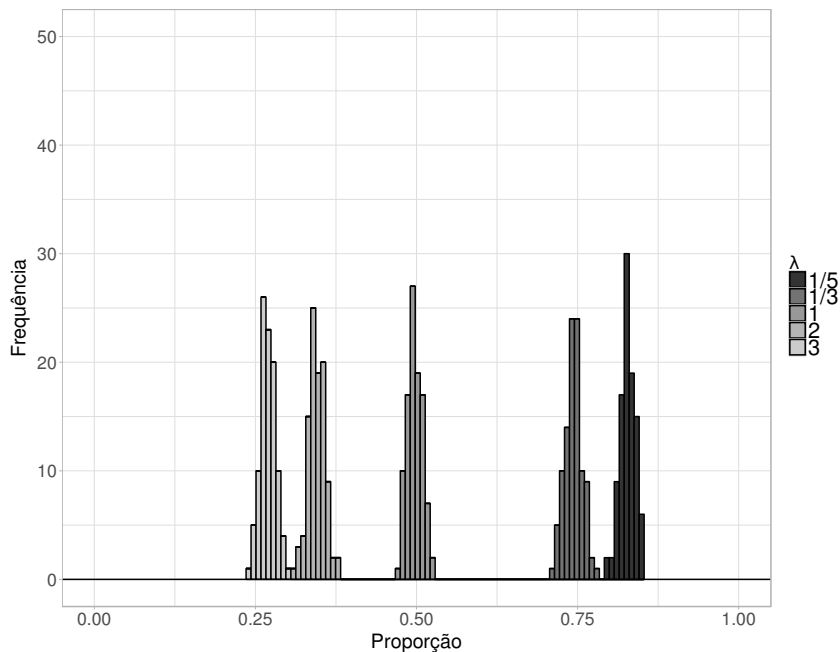


Figura 14 – Histograma das proporções de 1's nas 100 réplicas para cada cenário sob o modelo de regressão binária com ligação PL.

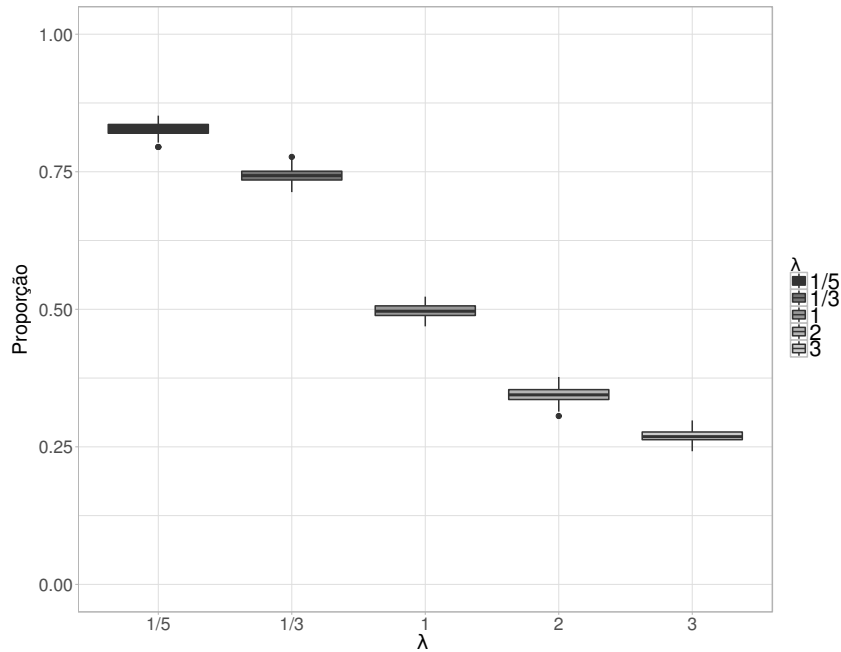


Figura 15 – Boxplot das proporções de 1's nas 100 réplicas para cada cenário sob o modelo de regressão binária com ligação PL.

Por outro lado, nas Figuras 16 e 17 pode-se observar que para $0 < \lambda < 1$ o histograma e o boxplot apresentam valores menores das proporções de 1's. Como esperado, quando $\lambda = 1$ é observado o histograma e o boxplot com valores das proporções de 1's próximos a 0,5. Por sua vez, no caso dos valores de $\lambda > 1$ o histograma e o boxplot apresentam valores maiores das proporções de 1's.

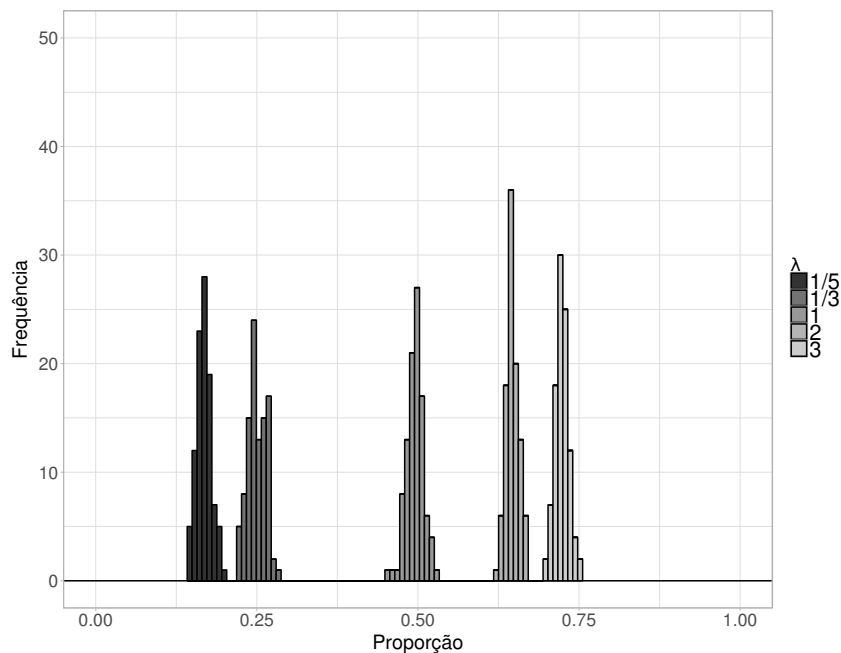


Figura 16 – Histograma das proporções de 1's nas 100 réplicas para cada cenário sob o modelo de regressão binária com ligação RPL.

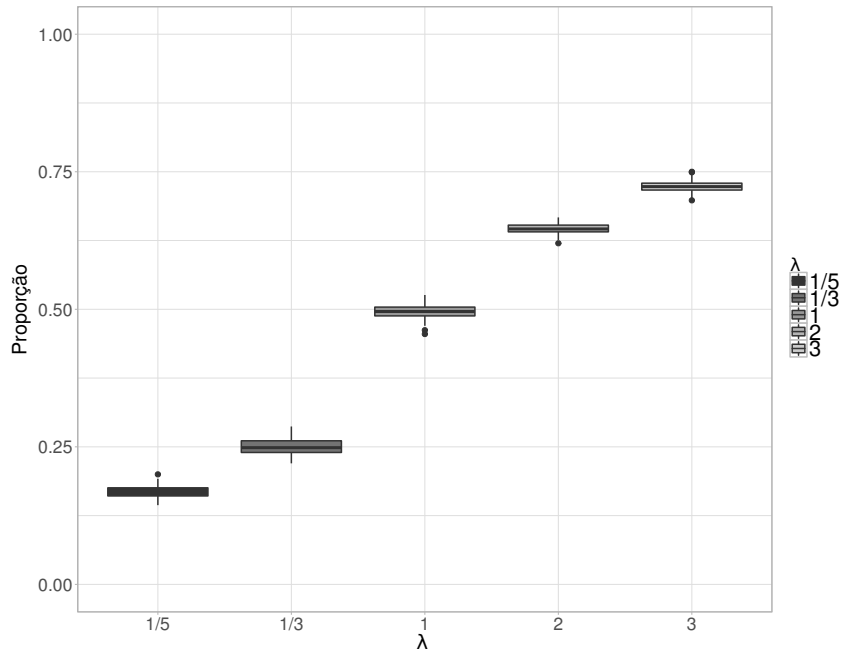


Figura 17 – Boxplot das proporções de 1's nas 100 réplicas para cada cenário sob o modelo de regressão binária com ligação RPL.

Semelhantes conclusões são obtidas ao simular dados dos modelos com as outras funções de ligação potência ou reversa de potência.

2.6 Interpretação do parâmetro λ

O parâmetro λ pode ser interpretado segundo a relação que a distribuição potência ou reversa de potência tem com a distribuição base usada para construí-la.

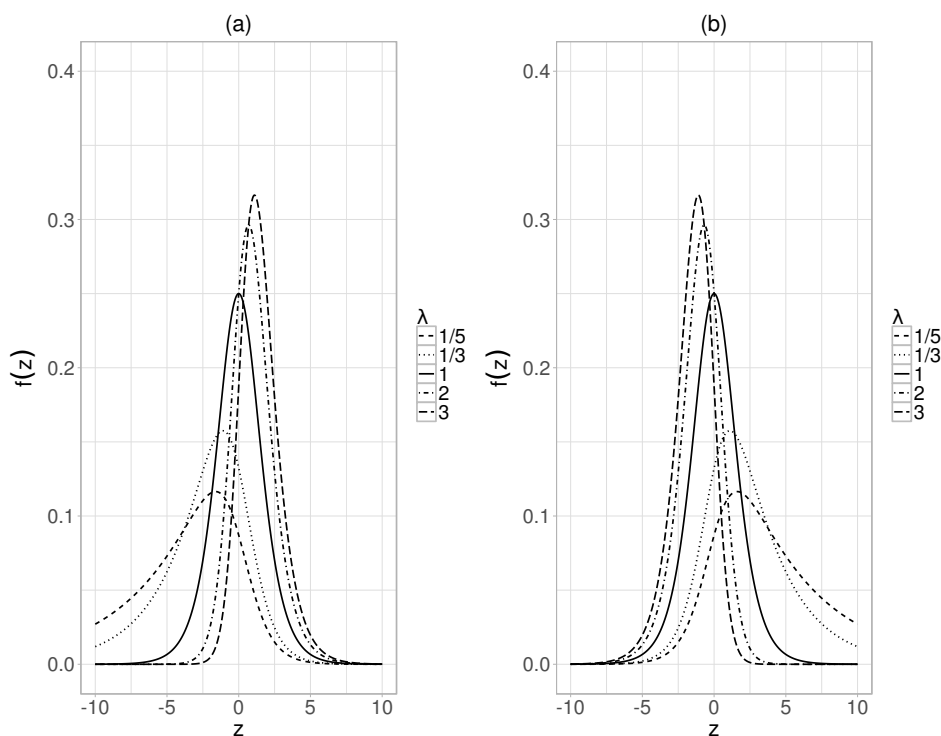


Figura 18 – Fdp da distribuição (a) Potência Logística e (b) Reversa de Potência Logística.

A Figura 18 mostra diferentes curvas da fdp da distribuição PL e da distribuição PLR para $\lambda = 1/5; 1/3; 1; 2; 3$. Observa-se que quando $\lambda = 1$ as curvas das duas distribuições correspondem à fdp da Logística. Quando $\lambda > 1$ as curvas das fdp distribuições Potência Logística apresentam deslocamento à direita e as curvas das fdp Reversa de Potência Logística apresentam deslocamento à esquerda. Também, nesses valores de $\lambda > 1$, todas as fdp tem curtose maior do que a curva da distribuição Logística, que é a distribuição base. Enquanto que quando o valor de $0 < \lambda < 1$, as curvas das fdp distribuições Potência Logística apresentam deslocamento à esquerda e as curvas das fdp Reversa de Potência Logística apresentam deslocamento à direita. Nesses valores de λ , todas as fdp tem curtose menor do que a curva da distribuição Logística. O mesmo comportamento pode ser observado para as distribuições potência e reversa de potência dos outros casos estudados neste trabalho.

Pode ser visto também que a distribuição Potência Logística tem densidade unimodal que é assimétrica à direita se $\lambda > 1$ e é assimétrica à esquerda se $0 < \lambda < 1$. Considerando a Reversa de Potência Logística cuja densidade é também unimodal, ela é assimétrica à esquerda se $\lambda > 1$ e é assimétrica à direita se $0 < \lambda < 1$. Para estender estas relações estabelecidas visualmente, um coeficiente de assimetria apropriado será introduzido na seguinte seção.

Ao seguir será estendida a descrição da influência de λ na variação da probabilidade de sucesso p e a análise da assimetria nas distribuições potência e reversa de potência.

Influência de λ na variação da probabilidade de sucesso

Esta seção é dedicada aos detalhes da influência do parâmetro λ na variação da probabilidade de sucesso p .

A Figura 19 mostra os gráficos das funções da probabilidade de sucesso $p = F_{\lambda}(\cdot)$ segundo o parâmetro λ nas distribuições potência e reversa de potência, respectivamente, com distribuição base Logística. Para o preditor linear, consideram-se valores fixos $\eta = -2.5; 0; 2.5$ pois ao considerar diferentes valores de η , diferentes curvas são obtidas. Também valores de $0 < \lambda \leq 10$ são considerados nos gráficos.

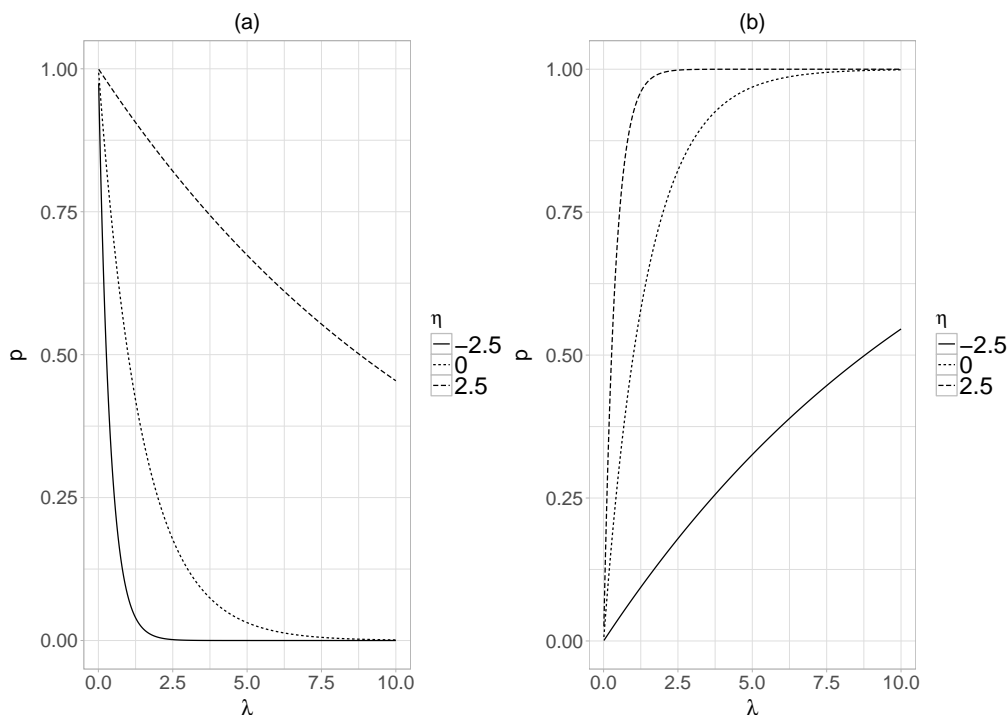


Figura 19 – Função da probabilidade de sucesso p segundo o parâmetro λ para $\eta = -2.5; 0; 2.5$ nas distribuições (a) Potência Logística e (b) Reversa de Potência Logística.

No caso da PL, na Figura 19 (a), pode ser visto que as curvas para todos os valores de η são decrescentes, isto quer dizer que, conforme λ aumenta, menor é a probabilidade p . A diminuição no valor de p é maior quando λ tiver valores próximos ao 0. Além disso, o valor da mudança em p devido ao incremento em λ é diferente dependendo do valor de η . Se η for negativo as variações na probabilidade p serão maiores.

No caso da PLR, na Figura 19 (b), pode ser visto que as curvas para todos os valores de η são crescentes, isto quer dizer que, conforme λ aumenta, maior é a probabilidade p . O incremento no valor de p é maior quando λ tiver valores próximos ao 0. Além do mais, o valor

da mudança em p devido ao incremento em λ é diferente dependendo do valor de η . Se η for negativo as variações na probabilidade p serão menores.

Para analisar as mudanças observadas nas figuras anteriores, define-se a derivada de p com respeito a λ que será obtida para cada caso.

No caso das distribuições potência, a derivada é:

$$\frac{\partial p}{\partial \lambda} = \frac{\partial F_P(z)}{\partial \lambda} = \frac{\partial \left(G(z)^\lambda \right)}{\partial \lambda} = G(z)^\lambda \log G(z) \quad (2.11)$$

Pode ser visto que o valor de $\frac{\partial p}{\partial \lambda}$ nas distribuições potência será negativo sempre uma vez que o valor de $\log G(z) < 0$.

Na Figura 20 podem ser vistas as curvas dos valores da derivada $\frac{\partial p}{\partial \lambda}$ e dos valores de λ para $\eta = -2,5; 0; 2,5$.

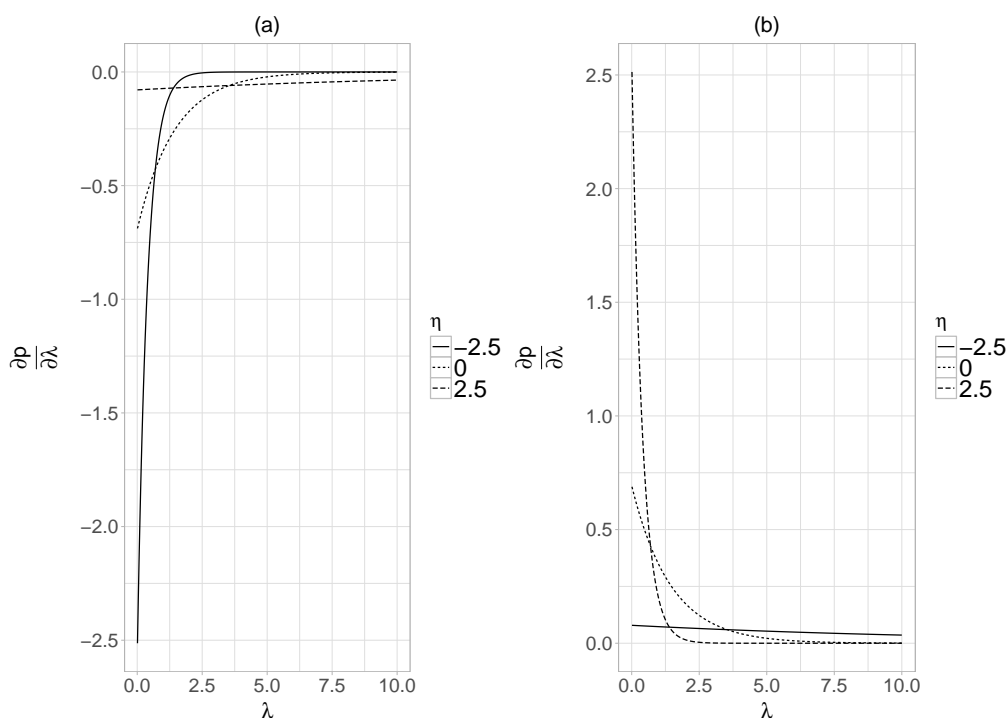


Figura 20 – Variação da probabilidade de sucesso p respeito ao parâmetro λ para $\eta = -2.5, 0, 2.5$ nas distribuições (a) Potência Logística e (b) Reversa de Potência Logística.

Na Figura 20 (a) pode ser observado que $\frac{\partial p}{\partial \lambda} < 0$ uma vez que as curvas da Figura 19 (a) foram todas decrescentes. Também que o valor da mudança em $\frac{\partial p}{\partial \lambda}$ devido ao incremento em λ é diferente dependendo do valor de η . Se η for menor as mudanças em $\frac{\partial p}{\partial \lambda}$ serão maiores. Outra observação é que para valores de λ que próximos de 0, as mudanças em $\frac{\partial p}{\partial \lambda}$ serão maiores.

No caso das distribuições reversa de potência, a derivada é:

$$\frac{\partial p}{\partial \lambda} = \frac{\partial F_{RP}(z)}{\partial \lambda} = \frac{\partial \left(1 - G(-z)^\lambda\right)}{\partial \lambda} = -G(-z)^\lambda \log G(-z) \quad (2.12)$$

É visto que o valor de $\frac{\partial p}{\partial \lambda}$ nas distribuições reversa de potência será positivo sempre uma vez que o valor de $\log G(-z) < 0$ e o sinal $-$ tornam o valor final positivo.

Na Figura 20 (b) pode ser observado que $\frac{\partial p}{\partial \lambda} > 0$ uma vez que as curvas da Figura 19 (b) foram todas crescentes. Também que o valor da mudança em $\frac{\partial p}{\partial \lambda}$ devido ao incremento em λ é diferente dependendo do valor de η . Se η for menor as mudanças em $\frac{\partial p}{\partial \lambda}$ serão menores. Outra observação é que para valores de λ que próximos de 0, as mudanças em $\frac{\partial p}{\partial \lambda}$ serão maiores.

Assimetria das distribuições potência e reversa de potência

O coeficiente de assimetria de Pearson (A_P) é definido como:

$$A_P = \frac{m_3(X)}{m_2(X)^{3/2}}, \quad (2.13)$$

onde os momentos são $m_3(X) = E[(X - \mu)^3]$ e $m_2(X) = E[(X - \mu)^2]$.

O calculo deste coeficiente para as distribuições potência e reversa de potência é complexo porque não existem formas fechadas para m_2 e m_3 (GUPTA; GUPTA, 2008).

Uma boa opção é o coeficiente de assimetria octil definido por Moors *et al.* (1996):

$$A_O = \frac{(O_7 - O_4) - (O_4 - O_1)}{O_7 - O_1}, \quad (2.14)$$

onde $0 < A_O < 1$, O_7 é o octil 7, O_4 o octil 4 e O_1 o octil 1. A Equação 2.10 também pode ser representada mediante os quantis:

$$A_O = \frac{(Q_{0,875} - Q_{0,5}) - (Q_{0,5} - Q_{0,125})}{Q_{0,875} - Q_{0,125}}. \quad (2.15)$$

Utilizando o coeficiente de assimetria octil, a assimetria dos modelos pode ser quantificada facilmente. Na Equação 2.11 é visto que são requeridos os quantis $Q_{0,875}$, $Q_{0,5}$ e $Q_{0,125}$, para obter-os precisa-se determina-os segundo as formas das distribuições potência e reversa de potência como mostrado a seguir.

Seja Q_p o quantil e $0 < p < 1$ a probabilidade a ser avaliada no quantil da distribuição de interesse. No caso das distribuições potência, o quantil obtido é:

$$G(q)^\lambda = p \Rightarrow Q_p = G^{-1}\left(p^{1/\lambda}\right). \quad (2.16)$$

No caso das distribuições reversa de potência, o quantil é dado a seguir:

$$1 - G(-q)^\lambda = p \Rightarrow Q_p = -G^{-1} \left((1-p)^{1/\lambda} \right). \quad (2.17)$$

Do anterior, é visto que o quantil para uma probabilidade p numa distribuição potência é:

$$Q_p^P = G^{-1} \left(p^{1/\lambda} \right)$$

Por sua vez, o quantil para uma probabilidade $1-p$ numa distribuição reversa de potência é:

$$Q_{1-p}^{RP} = -G^{-1} \left((1 - (1-p))^{1/\lambda} \right) = -G^{-1} \left(p^{1/\lambda} \right)$$

Assim, é possível estabelecer:

$$Q_{1-p}^{RP} = -Q_p^P \quad (2.18)$$

Na seguinte tabela os quantis para cada uma das distribuições potência e reversa de potência são apresentados.

Tabela 5 – Quantis para distribuições potência e reversa de potência.

Distribuição	Q^P	Q^{RP}
Logística	$Q_p^{PL} = \log \left(\frac{p^{1/\lambda}}{1-p^{1/\lambda}} \right)$	$Q_p^{RPL} = -\log \left(\frac{(1-p)^{1/\lambda}}{1-(1-p)^{1/\lambda}} \right)$
Normal	$Q_p^{PN} = \Phi^{-1} \left(p^{1/\lambda} \right)$	$Q_p^{RPN} = -\Phi^{-1} \left((1-p)^{1/\lambda} \right)$
Cauchy	$Q_p^{PC} = \tan \left(\pi \left(p^{1/\lambda} - 0,5 \right) \right)$	$Q_p^{RPC} = -\tan \left(\pi \left((1-p)^{1/\lambda} - 0,5 \right) \right)$
Gumbel de Valor Mínimo	$Q_p^{PGVMi} = \log \left(-\log \left(1 - p^{1/\lambda} \right) \right)$	$Q_p^{RPGVMi} = -\log \left(-\log \left(1 - (1-p)^{1/\lambda} \right) \right)$
Gumbel de Valor Máximo	$Q_p^{PGVMa} = -\log \left(-\log \left(p^{1/\lambda} \right) \right)$	$Q_p^{RPGVMa} = \log \left(-\log \left((1-p)^{1/\lambda} \right) \right)$

Considerando as distribuições potência, o coeficiente de assimetria octil terá a seguinte forma:

$$A_O^P = \frac{Q_{0,875}^P - 2Q_{0,5}^P + Q_{0,125}^P}{Q_{0,875}^P - Q_{0,125}^P}$$

Para as distribuições reversa de potência o coeficiente de assimetria octil será:

$$A_O^{RP} = \frac{Q_{0,875}^{RP} - 2Q_{0,5}^{RP} + Q_{0,125}^{RP}}{Q_{0,875}^{RP} - Q_{0,125}^{RP}}$$

Segundo a Equação 2.14, o coeficiente de assimetria octil nas distribuições reversa de potência pode-se expressar da seguinte forma.

$$A_O^{RP} = \frac{-Q_{0,125}^P + 2Q_{0,5}^P - Q_{0,875}^P}{-Q_{0,125}^P + Q_{0,875}^P} = \frac{- \left(Q_{0,875}^P - 2Q_{0,5}^P + Q_{0,125}^P \right)}{Q_{0,875}^P - Q_{0,125}^P}$$

Dessa forma é possível estabelecer a seguinte relação entre os coeficientes de assimetria octil das distribuições potência e reversa de potência:

$$A_O^{RP} = -A_O^P \quad (2.19)$$

Na Tabela 6 são apresentados os valores de A_O para cada distribuição.

Tabela 6 – Valores da assimetria octil para para distribuições potência e reversa de potência.

Distribuição	A_O			
	Mínimo	Máximo	$0 < \lambda < 1$	$\lambda \geq 1$
PL	-0,425	0,178	[-0,425; 0)	[0; 0,178]
RPL	-0,178	0,425	[0,425; 0)	[0; -0,178]
PN	-0,122	0,080	[-0,122; 0)	[0; 0,080]
RPN	-0,080	0,122	[0,122; 0)	[0; -0,080]
PC	-1	0,716	[-1; 0)	[0; 0,716]
RPC	-0,716	1	[1; 0)	[0; -0,716]
PGVMi	-0,422	-0,011	[-0,422; -0,199)	[-0,199; -0,011]
RPGVMi	0,011	0,422	[0,422; 0,199)	[0,199; 0,011]
PGVMa	0,199	0,199	0,199	0,199
RPGVMa	-0,199	-0,199	-0,199	-0,199

Nas Figuras 21 - 25 são apresentados os gráficos dos valores de A_O segundo diferentes valores de λ para cada uma das distribuições potência e reversa de potência usadas como funções de ligação e seguem o comportamento descrito na Tabela 6. Visualmente pode ser verificada a relação da Equação 2.15.

Note que os valores de assimetria das distribuições são diferentes. Na Figura 23, pode ser visto que as distribuições PC e RPC apresentam maior assimetria. Por sua vez, na Figura 22 pode ser visto que as distribuições PN e RPN apresentam menor assimetria. Note também, na Figura 25, que nas distribuições PGVMa e RPGMa a assimetria é constante e não depende do valor de λ , esse resultado será explicado proximamente.

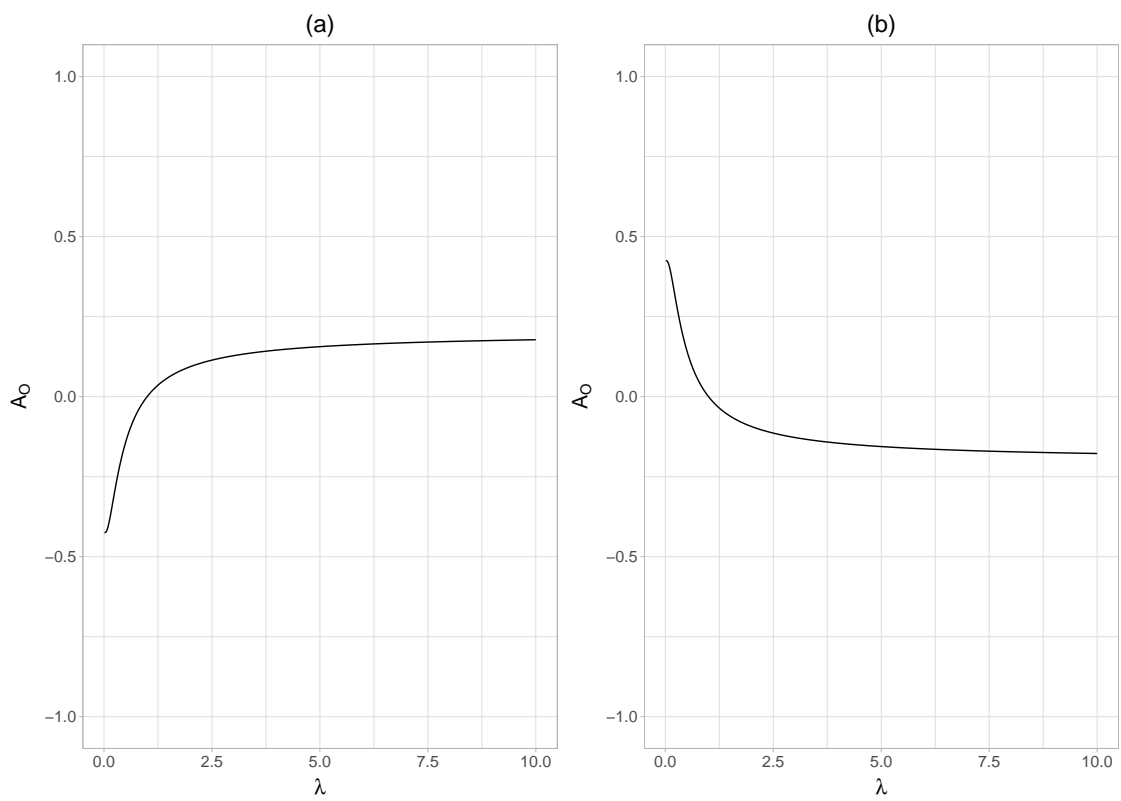


Figura 21 – Assimetria octil nas distribuições (a) PL e (b) RPL.

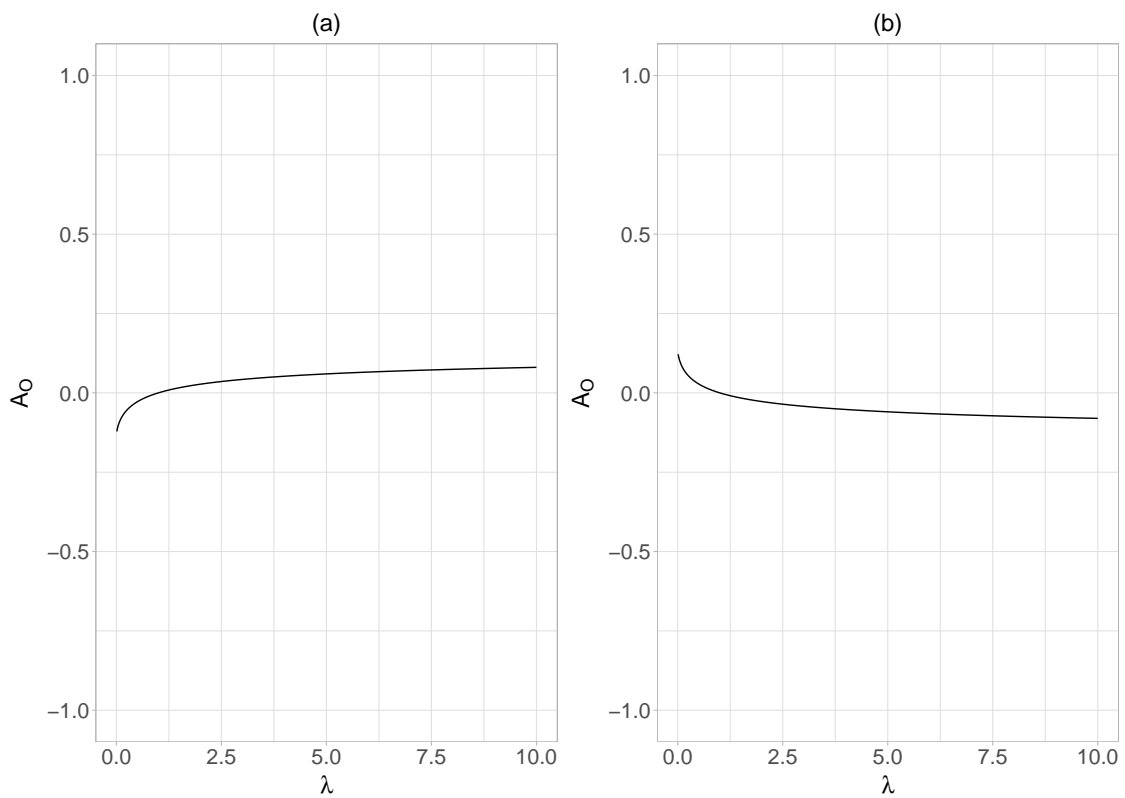


Figura 22 – Assimetria octil nas distribuições (a) PN e (b) RPN.

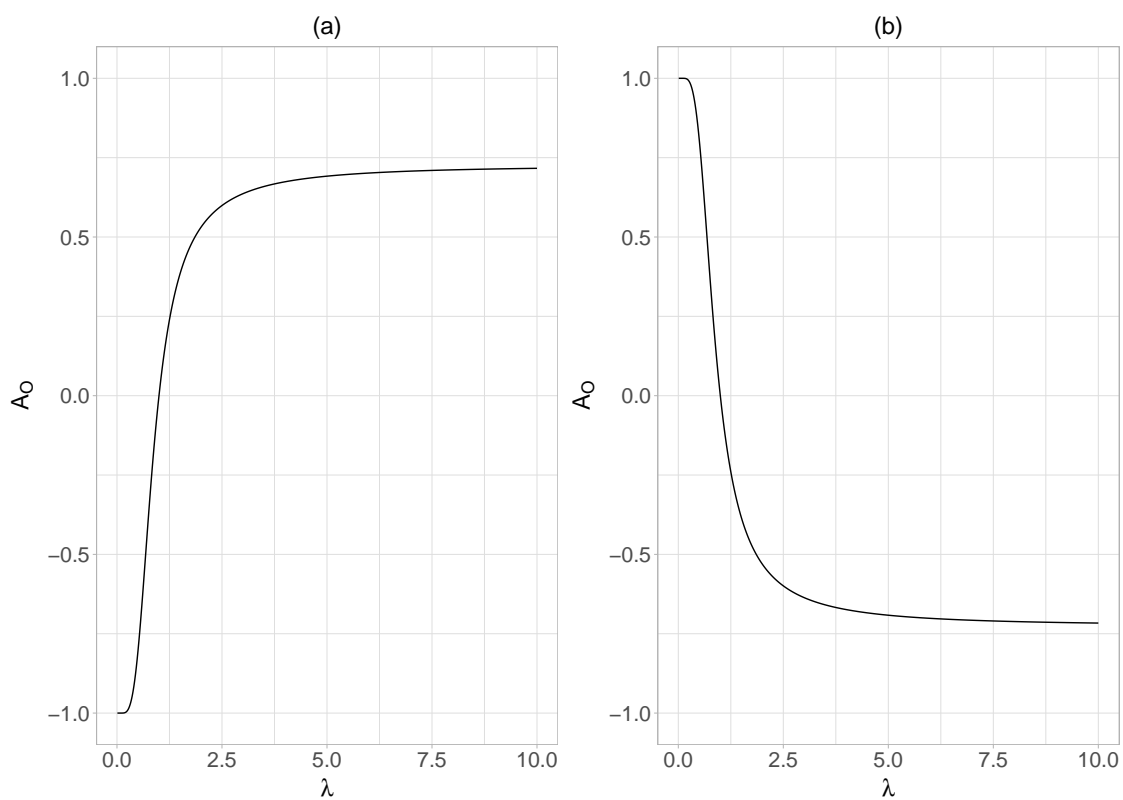


Figura 23 – Assimetria octil nas distribuições (a) PC e (b) RPC.

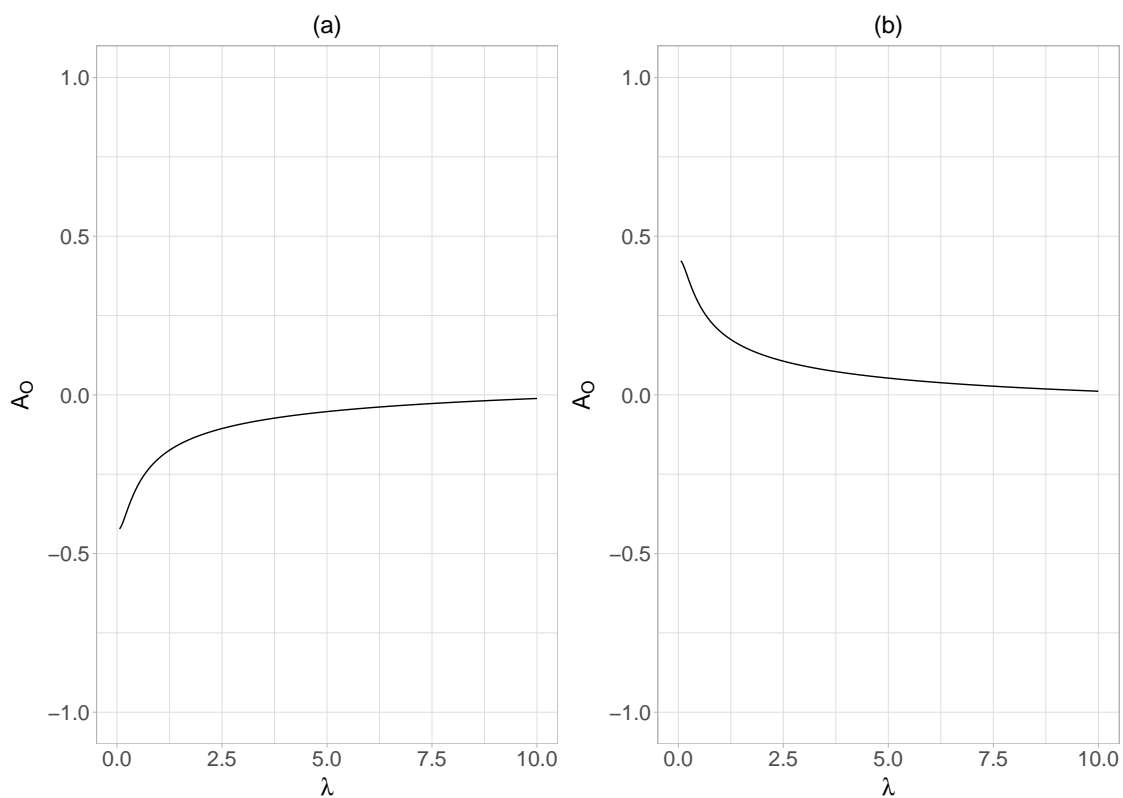


Figura 24 – Assimetria octil nas distribuições (a) PGVMi e (b) RPGVMi.

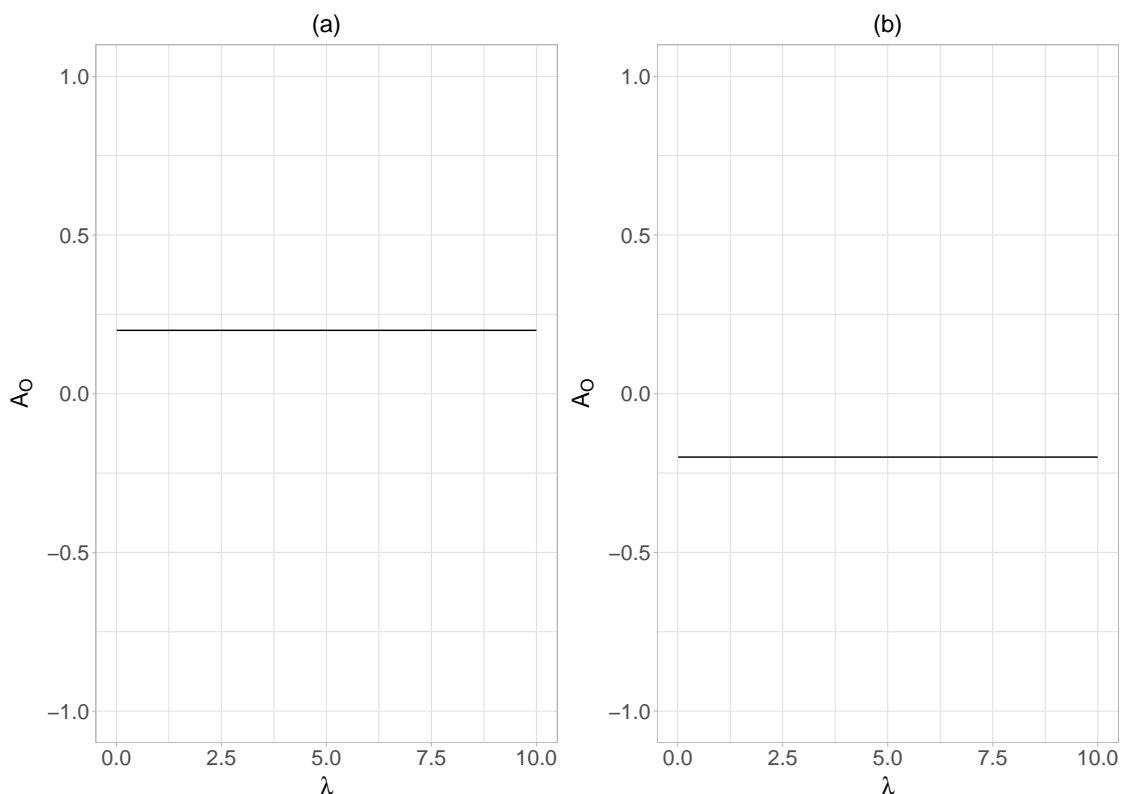


Figura 25 – Assimetria octil nas distribuições (a) PGVMa e (b) RPGVMa.

Também, nas Figuras 21, 22 e 23, quando $\lambda = 1$, o valor de A_O é zero uma vez que as distribuições de base, Logística, Normal e Cauchy, são distribuições simétricas. Considerando a Figura 24, note que as distribuições PGVMi e RPGMi sempre são assimétricas, isto é, $\lambda = 1 \Rightarrow A_O \neq 0$.

Comentário.

Observa-se na Figura 25 que no caso da distribuição Potência Gumbel de Valor Máximo, $A_O = 0,199$ e na distribuição Reversa de Potência Gumbel de Valor Máximo, $A_O = -0,199$, de forma constante para qualquer valor de λ . Isso acontece porque o efeito do parâmetro λ nessas distribuições traslada as curvas das distribuições horizontalmente, mantendo a forma original da distribuição base.

Na Figura 26 são observadas as curvas das funções de densidade para a distribuição Potência Gumbel de Valor Máximo com parâmetro $\lambda = 1/5; 1/3; 1; 2; 3$ trasladando-se. Nas Figuras 12 e 13 tinha-se visto também as curvas das suas funções de distribuição acumulada para os mesmos valores do parâmetro λ sem mudanças tão evidentes quanto as apresentadas nas outras distribuições.

A traslação horizontal das curvas, ou seja, existindo um acréscimo ou diminuição nos valores dos quantis é explicada a seguir.

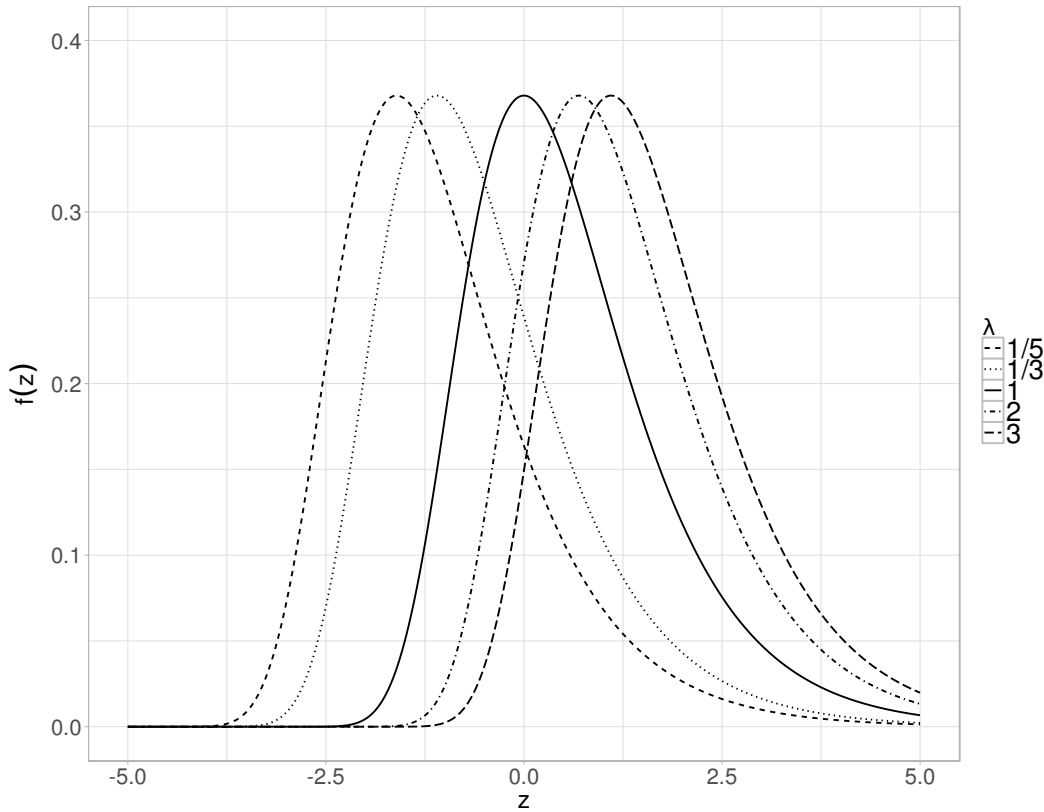


Figura 26 – Densidade da distribuição Potência Gumbel de Valor Máximo.

Considere-se $p^* = \int_{-\infty}^{q^*} f_{PGVMA}(z) dz$, onde p^* é a probabilidade acumulada de $f_{PGVMA}(z)$ desde $-\infty$ até algum quantil q^* . Porém, segundo a Tabela 5, tem-se que o quantil da distribuição vai ter a forma a seguir $q^* = -\log\left(-\log\left(p^{*1/\lambda}\right)\right)$. Considerando isto, logo pode-se partir da forma do quantil q^* para encontrar o valor que corresponde segundo $f_{PGVMA}(\cdot)$, fazendo a reposição q^* dentro dessa função.

$$\begin{aligned}
 f_{PGVMA}(q^*) &= \lambda \left[e^{-(q^*+e^{-q^*})} \right] \left[e^{-e^{-q^*}} \right]^{\lambda-1} = \lambda \left[e^{-q^*} \right] \left[e^{-e^{-q^*}} \right] \left[e^{-e^{-q^*}} \right]^{\lambda-1} \\
 &= \lambda \left[e^{-q^*} \right] \left[e^{-e^{-q^*}} \right]^{\lambda} = \lambda \left[e^{\log(-\log(p^{*1/\lambda}))} \right] \left[e^{-e^{\log(-\log(p^{*1/\lambda}))}} \right]^{\lambda} \\
 &= \lambda \left[-\log\left(p^{*1/\lambda}\right) \right] \left[e^{\log(p^{*1/\lambda})} \right]^{\lambda} = \lambda \left[-\frac{1}{\lambda} \log(p^*) \right] \left[p^{*1/\lambda} \right]^{\lambda} \\
 &= \left[-\log(p^*) \right] \left[p^* \right]
 \end{aligned}$$

O valor final é sempre positivo uma vez que os valores negativos do logaritmo de p^* são tomados, ademais a expressão final só depende do valor de p^* e não de λ . Como tem sido tomados valores arbitrários para p^* e q^* , pode-se estender a ideia de que seja qualquer o valor do $\lambda > 0$, $f_{PGVMA}(q^*)$ não depende de λ . Pode-se ver que λ só muda o valor dos quantis como q^* . Isto quer dizer que se foram tomados diferentes valores de λ , os quantis mudarão, porém o valor correspondente de cada quantil na densidade de $f_{PGVMA}(z)$ é independente de λ . Pode-se concluir que existe a translação horizontal nas densidades.

Além da explicação anterior, na Figura 27 pode ser vista a curva das probabilidades e os quantis da distribuição PGVMa com parâmetro $\lambda = 1; 2$. Se alguma probabilidade de referência p nas duas curvas for tomada, a diferença nos valores dos quantis é mantida constante. No caso apresentado na figura, a diferença nos quantis terá o valor de 0,6931, independentemente do valor de p escolhido e sinalizado com linhas para $p = 0,125; 0,5; 0,875$.

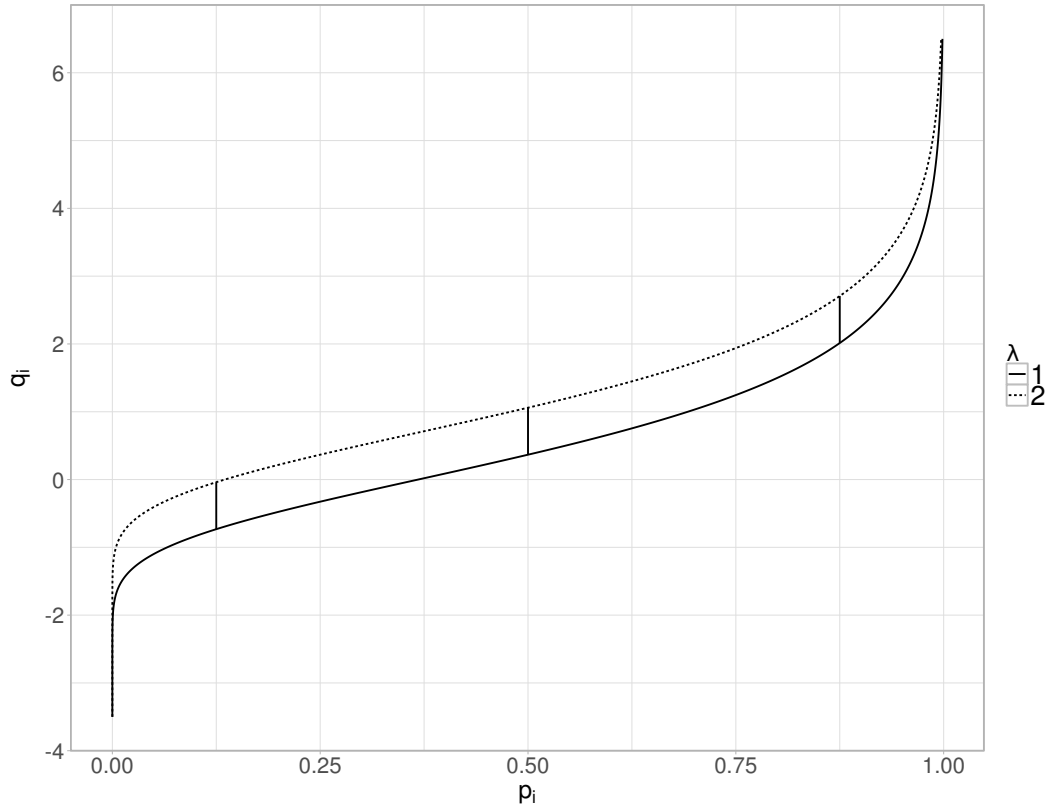


Figura 27 – Probabilidades e quantis da distribuição Potência Gumbel de Valor Máximo.

Para entender esta característica, considere duas distribuições PGVMa com λ_1 e λ_2 . Seja κ a diferença nos quantis das distribuições:

$$Q_p^{\text{PGMa}}_{\lambda_1} - Q_p^{\text{PGMa}}_{\lambda_2} = \kappa$$

$$\kappa = -\log\left(-\log\left(p^{1/\lambda_1}\right)\right) - -\log\left(-\log\left(p^{1/\lambda_2}\right)\right) = \log\left(-\log\left(p^{1/\lambda_2}\right)\right) - \log\left(-\log\left(p^{1/\lambda_1}\right)\right)$$

$$\kappa = \log\left(\frac{\log\left(p^{1/\lambda_2}\right)}{\log\left(p^{1/\lambda_1}\right)}\right) \Rightarrow e^\kappa = \frac{\log\left(p^{1/\lambda_2}\right)}{\log\left(p^{1/\lambda_1}\right)} = \frac{\frac{1}{\lambda_2}}{\frac{1}{\lambda_1}} = \frac{\lambda_1}{\lambda_2}$$

$$\kappa = \log\left(\frac{\lambda_1}{\lambda_2}\right)$$

O valor de κ só depende dos possíveis valores de λ e não dos valores da potência de $1/\lambda$ em p . O mesmo acontece para o caso da distribuição Reversa de Potência Gumbel de Valor Máximo. Porém, se essa diferença κ entre quantis das outras distribuições potência ou reversa de potência for calculada, pode-se ver que κ dependerá de dos possíveis valores da potência de $1/\lambda$ em p , ou seja, a existência de $1/\lambda$ como potência de p será manifestada quantitativamente. Consequentemente as curvas das distribuições serão diferentes e a traslação não poderia ser observada.

Para o caso na Figura 27 onde considera-se a distribuição Potência Gumbel de Valor Máximo com parâmetro $\lambda = 1;2$, $\kappa = \log\left(\frac{2}{1}\right) = 0,6931$. Esta característica pode ser simplesmente verificada com quaisquer valores de λ .

Interpretação de λ para modelos de regressão binária.

- Como visto na subseção 2.5, nos modelos de regressão binária com funções de ligação potência, o valor de λ está relacionado ao valor da proporção de 1's na variável resposta. Assim, para modelos onde $0 < \lambda < 1$, a proporção de 1's na variável resposta é maior a 0,5, ou seja, existem mais sucessos do que falhas. Para $\lambda = 1$ a proporção de 1's é próxima a 0,5, ou seja, existe balanceamento entre os sucessos e as falhas. Para $\lambda > 1$ a proporção de 1's é menor a 0,5, ou seja, existem mais falhas do que sucessos. Segundo o visto na subseção 2.6.2, quando consideram-se modelos de regressão binária com ligações potência logit, probit ou cachit; quando $0 < \lambda < 1$ a assimetria octil será negativa, quando $\lambda = 1$ a assimetria será 0 (ou simétrica), quando $\lambda > 1$ a assimetria octil será positiva. Quando considera-se a ligação cloglog a assimetria octil será sempre negativa para qualquer valor de $\lambda > 0$. Quando considera-se a ligação loglog, a assimetria octil será constante com valor 0,199 para qualquer valor de $\lambda > 0$.
- Por sua vez, nos modelos de regressão binária com funções de ligação reversa de potência, o valor de λ está também relacionado ao valor da proporção de 1's na variável resposta. Assim, para modelos onde $0 < \lambda < 1$, a proporção de 1's na variável resposta é menor a 0,5, ou seja, existem mais falhas do que sucessos. Para $\lambda = 1$ a proporção de 1's é próxima a 0,5, ou seja, existe balanceamento entre os sucessos e as falhas. Para $\lambda > 1$ a proporção de 1's é maior a 0,5, ou seja, existem mais sucessos do que falhas. Por sua vez, respeito à assimetria, quando consideram-se modelos de regressão binária com ligações reversa de potência logit, probit ou cachit; quando $0 < \lambda < 1$ a assimetria octil será positiva, quando $\lambda = 1$ a assimetria será 0 (ou simétrica), quando $\lambda > 1$ a assimetria octil será negativa. Quando considera-se a ligação potência cloglog a assimetria octil será sempre positiva para qualquer valor de $\lambda > 0$. Quando considera-se a ligação potência loglog, a assimetria octil será constante com valor -0,199 para qualquer valor de $\lambda > 0$.

ESTIMAÇÃO E AJUSTE DOS MODELOS

Neste Capítulo são abordados os detalhes da estimação bayesiana dos modelos propostos no Capítulo 2, como a especificação das distribuições a priori, a descrição dos algoritmos Metropolis-Hastings dentro de Gibbs e No-U-Turn Sampler. Também são dadas as definições das medidas de comparação de modelos segundo os critérios de informação como DIC, EAIC, EBIC, WAIC e LOO, e medidas de avaliação preditiva como o AUC; que serão úteis na escolha de modelos apropriados

3.1 Estimação bayesiana

Em geral para qualquer modelo de regressão binária com função de ligação potência ou reversa de potência apresentado na seção 2.4, a verossimilhança é dada por:

$$\mathbf{L}(\boldsymbol{\beta}, \lambda \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n [F_{\lambda}(\eta_i)]^{y_i} [1 - F_{\lambda}(\eta_i)]^{1-y_i}. \quad (3.1)$$

Onde F_{λ} pode ser quaisquer das fda das distribuições potência ou reversa de potência introduzidas no Capítulo 2. Especificamente consideramos a distribuição potência logit, reversa de potência logit, potência probit, reversa de potência probit, potência cloglog, reversa de potência cloglog, potência loglog, reversa de potência loglog, potência cauchit ou reversa de potência cauchit, apresentadas na Tabela 4. Na verossimilhança dada acima, $F_{\lambda}(\eta_i)$ representa a probabilidade de sucesso na distribuição de Bernoulli da variável resposta y_i , isto é $P(Y_i = 1) = F_{\lambda}(\eta_i)$.

Note que neste tipo de modelo os parâmetros $\boldsymbol{\beta}$ e λ têm um significado diferente. Por um lado, λ é um parâmetro estrutural associado com a escolha da ligação. Por outro lado, $\boldsymbol{\beta}$ é um vetor de parâmetros estruturais inerentes às observações e não depende da escolha do modelo. Então, dois casos de estimação podem ser considerados. No primeiro, os parâmetros

$\boldsymbol{\beta}$ e λ são conjuntamente estimados; no segundo, unicamente os parâmetros $\boldsymbol{\beta}$ são estimados e λ é assumido conhecido. O primeiro caso pode ser definido como estimação não condicional e o segundo como estimação condicional. Neste trabalho a estimação não condicional será desenvolvida, isto é, a estimação conjunta de ambos os tipos de parâmetros sob abordagem bayesiana. Por outro lado, a estimação condicional usando máxima verossimilhança perfilada seria uma possibilidade de trabalhos futuros.

Prioris

Segundo [Bazán, Romeo e Rodrigues \(2014\)](#), é conveniente introduzir a δ -parametrização, definindo $\delta = \log(\lambda)$, então uma distribuição a priori deve ser especificada para δ . Note que neste caso o parâmetro $\delta \in \mathbb{R}$. Considerando esta reparametrização do modelo, precisa-se especificar as distribuições a priori para $\boldsymbol{\beta}$ e δ , uma vez que ambos os parâmetros são de diferentes tipos; nós assumimos eles independentes, isto é:

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

Para $\boldsymbol{\beta}$, os coeficientes de regressão, assume-se $\beta_j \sim N(\mu_{\beta_j}, \sigma_{\beta_j}^2) \forall j = 0, \dots, k$. Neste caso considera-se como hiperparâmetros $\mu_{\beta_j} = 0$ e $\sigma_{\beta_j}^2 = 100$ denotando a ignorância em relação ao parâmetro. O anterior é uma prática comum em modelos de regressão como pode ser visto em [Jiang et al. \(2013\)](#). Para o parâmetro δ , seguindo a [Bazán, Romeo e Rodrigues \(2014\)](#) considera-se a *Uniforme* $(-2, 2)$ porque os valores fora do intervalo $[e^{-2}, e^2]$ não são observados empiricamente considerando esta especificação.

A estrutura hierárquica do modelo com a utilização da δ -parametrização é dada por:

$$\begin{aligned} Y_i | \boldsymbol{\beta}, \delta &\sim \text{Bernoulli}(p_i) \\ p_i &= F_\delta(x_i' \boldsymbol{\beta}), \quad i = 1, 2, \dots, n \\ \beta_j &\sim N(0, 100), \quad j = 0, 1, \dots, k \\ \delta &\sim U(-2, 2). \end{aligned} \tag{3.2}$$

Considerando a estrutura hierárquica dada em (3.2), a densidade da distribuição a posteriori dos modelos de regressão binária com função de ligação potência ou de potência, tem a seguinte forma:

$$\pi(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \propto \mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}) \pi(\delta)$$

$$\pi(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^n F_{\delta}(x'_i \boldsymbol{\beta})^{y_i} (1 - F_{\delta}(x'_i \boldsymbol{\beta}))^{1-y_i} \frac{1}{4} \prod_{j=0}^k \frac{1}{\sqrt{2\pi}\sqrt{100}} \exp\left\{-\frac{(\beta_j)^2}{2(100)}\right\} \quad (3.3)$$

Como a distribuição a posteriori não tem forma fechada ela pode ser simulada via algoritmos MCMC. Neste trabalho são considerados os algoritmos Metropolis-Hastings dentro de Gibbs e No-U-Turn Sampler para simular esta distribuição.

Estimação usando Metropolis-Hastings dentro de Gibbs

Na estimação bayesiana da regressão binária tem sido usado geralmente o algoritmo do amostrador de Gibbs (GEMAN; GEMAN, 1984), o qual é um caso especial do algoritmo Metropolis-Hastings desenvolvido no trabalho de Metropolis *et al.* (1953) e Hastings (1970). Neste trabalho é considerada a implementação do algoritmo Metropolis-Hastings Adaptativo dentro do amostrador de Gibbs (GILKS; BEST; TAN, 1995); o qual é um algoritmo que inclui a capacidade de aperfeiçoar uma função envelope passo a passo, do algoritmo adaptativo de rejeição (GILKS; WILD, 1992).

Nesta seção é apresentada a adaptação do algoritmo Metropolis-Hastings dentro de Gibbs para o caso da regressão binária com função de ligação potência logit. A verossimilhança é dada por:

$$\begin{aligned} \mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) &= \prod_{i=1}^n \left[F(\eta_i)^{\exp(\delta)} \right]^{y_i} \left[1 - F(\eta_i)^{\exp(\delta)} \right]^{1-y_i} \\ \mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) &= \prod_{i=1}^n \left[\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{y_i} \left[1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{1-y_i} \end{aligned}$$

A distribuição a posteriori para $(\boldsymbol{\beta}, \delta)$ é dada por:

$$\pi(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \propto \mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}, \delta)$$

$$\pi(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^n \left[\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{y_i} \left[1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{1-y_i} \frac{1}{4} I_{(-2,2)}(\delta) \prod_{j=0}^k \exp\left\{-\frac{\beta_j^2}{2\sigma_{\beta_j}^2}\right\}$$

No algoritmo de Metropolis-Hastings dentro de Gibbs, precisam-se as distribuições completas condicionais para os parâmetros de interesse $(\boldsymbol{\beta}, \lambda)$. São dadas por:

$$\pi(\beta_j | \boldsymbol{\beta}_{\setminus j}, \delta, \mathbf{y}, \mathbf{X}) = K_j^{-1} \prod_{i=1}^n \left[\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{y_i} \left[1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{1-y_i} \exp\left\{-\frac{\beta_j^2}{2\sigma_{\beta_j}^2}\right\},$$

$$\pi(\delta | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = K_d^{-1} \prod_{i=1}^n \left[\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{y_i} \left[1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right]^{1-y_i} \frac{1}{4} \mathbb{I}_{(-2,2)}(\delta),$$

onde K_j , $j = 0, 1, \dots, k$ e K_d são constantes normalizadoras.

Então tem-se:

- Para $t = 1, \dots, T$

1. Estabelecer $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{(t-1)}, \delta^{(t-1)} \right)^T$.

2. Para $j = 0, 1, \dots, k$

a) Propor um novo valor β'_j de $\pi(\beta_j | \boldsymbol{\beta}_{\setminus j}, \delta, \mathbf{y}, \mathbf{X})$.

b) Estabelecer $\boldsymbol{\theta}' = \left(\beta'_j, \boldsymbol{\beta}_{\setminus j}, \delta^{(t-1)} \right)^T$.

c) Calcular $\log \alpha = \min(0, A)$ onde A é dado por

$$A = \log \frac{\mathbf{L}(\beta'_j, \boldsymbol{\beta}_{\setminus j}, \delta | \mathbf{y}, \mathbf{X}) \pi(\beta'_j, \boldsymbol{\beta}_{\setminus j}, \delta)}{\mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}, \delta)}.$$

d) Atualizar $\boldsymbol{\theta} = \boldsymbol{\theta}'$ com probabilidade α ou manter os mesmos valores de $\boldsymbol{\theta}$ com probabilidade $1 - \alpha$.

3. Propor um novo valor δ' de $\pi(\delta | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$.

4. Estabelecer $\boldsymbol{\theta}' = \left(\boldsymbol{\beta}, \delta' \right)^T$.

5. Calcular $\log \alpha = \min(0, A)$ onde A é dado por

$$A = \log \frac{\mathbf{L}(\boldsymbol{\beta}, \delta' | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}, \delta')}{\mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}, \delta)}.$$

6. Atualizar $\boldsymbol{\theta} = \boldsymbol{\theta}'$ com probabilidade α ou manter os mesmos valores iniciais de $\boldsymbol{\theta}$ com probabilidade $1 - \alpha$.

7. Estabelecer $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$.

Dessa forma, o algoritmo que combina o Metropolis-Hasting dentro do amostrador de Gibbs pode ser implementado. A implementação será usando o pacote JAGS (PLUMMER, 2003)

Estimação usando o No-U-Turn Sampler

Como foi mencionado na introdução é de interesse fazer a estimação via algoritmo Monte Carlo Hamiltoniano (MCH) desenvolvido por [Duane et al. \(1987\)](#), na extensão No-U-Turn Sampler (NUTS) ([HOFFMAN; GELMAN, 2014](#)). A implementação do NUTS é feita no software Stan ([CARPENTER et al., 2016](#)). A seguir será apresentado o algoritmo para o caso da regressão binária com função de ligação potência logit.

Seja $\pi(\boldsymbol{\beta})$ a priori do vetor de parâmetros dos coeficientes de regressão $\boldsymbol{\beta}$, $\pi(\delta)$ a priori de δ e $\mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X})$ a verossimilhança. Sabe-se que existe a seguinte relação de proporcionalidade na distribuição a posteriori:

$$\pi(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \propto \mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}) \pi(\delta).$$

Segundo o trabalho de [Neal \(2012\)](#), a distribuição a posteriori pode-se representar como a distribuição canônica usando a energia potencial do sistema da dinâmica Hamiltoniana, $U(\mathbf{q})$, de modo que:

$$U(\boldsymbol{\beta}, \delta) = -\log[\mathbf{L}(\boldsymbol{\beta}, \delta | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}) \pi(\delta)]. \quad (3.4)$$

Pode ser visto que o vetor $(\boldsymbol{\beta}, \delta)$ faz o papel de vetor de posição \mathbf{q} na dinâmica Hamiltoniana. A forma de $U(\boldsymbol{\beta}, \delta)$ na equação (3.4) pode ser representada da seguinte forma:

$$U(\boldsymbol{\beta}, \delta) = \sum_{j=0}^k \frac{\beta_j^2}{2\sigma_{\beta_j}^2} - \log\left(\frac{1}{4}\right) I_{\delta}(-2, 2) - \sum_{i=1}^n \left[y_i \log \left[\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right] + (1 - y_i) \log \left[1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)} \right] \right].$$

Dessa forma depois devem ser obtidas as derivadas parciais da energia potencial, segundo cada elemento que compõe o vetor de posição:

$$\frac{\partial U}{\partial \beta_j} = \frac{\beta_j}{\sigma_{\beta_j}^2} - \sum_{i=1}^n \left[y_i \frac{\exp(\delta) \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{(\exp(\delta)-1)} \left(\frac{x_{ij} \exp(\eta_i)}{(1 + \exp(\eta_i))^2} \right)}{\left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)}} + (1 - y_i) \frac{-\exp(\delta) \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{(\exp(\delta)-1)} \left(\frac{x_{ij} \exp(\eta_i)}{(1 + \exp(\eta_i))^2} \right)}{1 - \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)^{\exp(\delta)}} \right],$$

$$\frac{\partial U}{\partial \delta} = - \sum_{i=1}^n \left[y_i \log \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right) \exp(\delta) - (1 - y_i) \log \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right) \frac{\exp(\eta_i) \exp(\delta)}{(1 + \exp(\eta_i))^{\exp(\delta)} - \exp(\eta_i) \exp(\delta)} \exp(\delta) \right].$$

O conjunto de derivadas parciais obtidas para cada uma das variáveis da energia potencial, $\frac{\partial U}{\partial \boldsymbol{\beta}}$ e $\frac{\partial U}{\partial \delta}$, terá a notação de $\nabla_{\mathbf{q}} U(\mathbf{q})$. Como detalhado anteriormente, no modelo de regressão binária com função de ligação potência logit o vetor de posição $\mathbf{q} = \boldsymbol{\theta} = (\boldsymbol{\beta}, \delta)$, é de $k + 2$ dimensões

($d = k + 2$). Dessa forma, o vetor de momento \mathbf{r} também será de $k + 2$ dimensões, com valores gerados aleatoriamente da distribuição Normal multivariada de média zero com os componentes de \mathbf{r} independentes, onde cada i tem variância m_i . De forma que:

$$K(\mathbf{r}) = \sum_{i=1}^d \frac{r_i^2}{2m_i}.$$

Com a distribuição canônica da energia potencial, $U(\mathbf{q})$, e as derivadas parciais, $\nabla_{\mathbf{q}}U(\mathbf{q})$; pode-se fornecer o algoritmo do MCH (NEAL, 2012).

- Para $t = 1, \dots, T$
 1. Estabelecer o vetor de posição inicial $\mathbf{q} = \mathbf{q}^{(t-1)}$.
 2. Gerar os d valores iniciais do vetor de momento \mathbf{r} .
 3. Estabelecer $\mathbf{r}^{(t-1)} = \mathbf{r}$.
 4. Fazer $\mathbf{r}' = \mathbf{r} - (\varepsilon/2)\nabla_{\mathbf{q}}U(\mathbf{q})$.
 5. De $i = 1$ até L :
 - a) Fazer $\mathbf{q}' = \mathbf{q} + \varepsilon\mathbf{r}'$.
 - b) Se $i = L$, fazer $\mathbf{r}' = \varepsilon\nabla_{\mathbf{q}}U(\mathbf{q}')$.
 6. Fazer $\mathbf{r}' = \mathbf{r}' - (\varepsilon/2)\nabla_{\mathbf{q}}U(\mathbf{q}')$.
 7. Fazer $\mathbf{r}' = -\mathbf{r}'$.
 8. Calcular $\alpha = \min(1, A)$ onde A é dado por

$$A = \exp\left(U(\mathbf{q}) - U(\mathbf{q}') + K(\mathbf{r}) - K(\mathbf{r}')\right)$$

9. Gerar u de $U(0, 1)$. Se $u < \alpha$, $\mathbf{q}^{(t)} = \mathbf{q}'$, caso contrario, $\mathbf{q}^{(t)} = \mathbf{q}^{(t-1)}$

No algoritmo MCH, precisa-se da discretização do tempo para obter a aproximação das equações hamiltonianas de modo que o método possa ser implementado no computador. Para poder fazer tal discretização utiliza-se o método Leapfrog, que é implementado no quinto passo do algoritmo dado acima. No método Leapfrog é necessário calibrar os parâmetros ε e L que determinam o tamanho do passo na discretização e o número de passos, respectivamente. No trabalho, as simulações e aplicações serão feitas com o algoritmo No-U-Turn, uma forma adaptativa do MCH que elimina a necessidade da calibração dos parâmetros da discretização. Isto é importante pois escolher valores pequenos ou grandes desses parâmetros pode trazer possíveis problemas como os detalhados na Tabela 7.

Para maiores detalhes sobre o MCH recomenda-se revisar o Anexo A.

Tabela 7 – Possíveis problemas na escolha dos parâmetros do método Leapfrog (tabela elaborada do trabalho de Neal (2012).)

Parâmetros	valor pequeno	valor grande
ε	gasto computacional	simulação imprecisa e alta taxa de rejeição
L	comportamento semelhante ao random walk	as trajetórias retraçam o passo

O algoritmo NUTS começa fazendo a re-amostragem no vetor do momento do sistema, r . Depois disso o algoritmo faz amostragem da variável u que segue distribuição uniforme e transforma a distribuição condicional dos parâmetros de posição numa distribuição uniforme, o que contribui a implementação do NUTS. O algoritmo constrói uma árvore binária com nós que representam a posição e o momento do sistema, na construção da árvore, no passo j são construídos $2j - 1$ nós novos e escolhe-se aleatoriamente a direção da sequência dos nós novos (HOFFMAN; GELMAN, 2014).

3.2 Comparação de modelos

A estrutura dos modelos a serem desenvolvidos foi descrita nas seções anteriores. Quando têm-se vários modelos as comparações entre eles podem ser feitas segundo o descrito em Gelman, Hwang e Vehtari (2014). Alguns dos critérios de comparação são descritos ao seguir.

Critérios de informação

Considerando $s = 1, \dots, S$ valores da amostra válida, obtidos usando os métodos de estimação descritos na seção 3.1 podem ser definidos os seguintes critérios.

DIC

O Deviance Information Criterion (DIC) foi proposto por Spiegelhalter *et al.* (2002) e é baseado na média a posteriori do desvio, $E[D(\boldsymbol{\beta}, \boldsymbol{\delta})]$ que computacionalmente é aproximada por:

$$Dbar = \frac{1}{S} \sum_{s=1}^S D(\boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)}), \quad (3.5)$$

para $s = 1, \dots, S$ onde S é o tamanho da amostra a posteriori.

O desvio é duas vezes o logaritmo da função de verossimilhança dadas as estimativas pontuais do modelo ajustado, ou seja:

$$D(\boldsymbol{\beta}, \boldsymbol{\delta}) = -2 \log(p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta})) = -2 \sum_{i=1}^n \log p(Y_i | \boldsymbol{\beta}, \boldsymbol{\delta}), \quad (3.6)$$

Para obter o DIC precisa-se de ρ_D , o número efetivo de parâmetros, aproximado computacionalmente por:

$$\widehat{\rho}_D = E \left[D(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\delta}}) \right] - D \left[E(\widehat{\boldsymbol{\beta}}), E(\widehat{\boldsymbol{\delta}}) \right] = Dbar - Dhat, \quad (3.7)$$

onde $Dhat = D \left[\widehat{E}(\boldsymbol{\beta}), \widehat{E}(\boldsymbol{\delta}) \right]$ é o desvio da média a posteriori, que é obtido das médias dos valores gerados da distribuição a posteriori:

$$Dhat = D \left(\frac{1}{S} \sum_{s=1}^S \boldsymbol{\beta}^{(s)}, \frac{1}{S} \sum_{s=1}^S \boldsymbol{\delta}^{(s)} \right). \quad (3.8)$$

A estimativa do DIC é dada por:

$$\widehat{DIC} = Dbar + \widehat{\rho}_D = 2Dbar - Dhat. \quad (3.9)$$

EAIC

O Expected Akaike Information Criterion (EAIC) foi apresentado em [Brooks \(2002\)](#) e é dado ao seguir:

$$\widehat{EAIC} = Dbar + 2v. \quad (3.10)$$

Onde v é o número de parâmetros.

EBIC

O Expected Bayesian Information Criterion (EBIC) foi apresentado em [Brooks \(2002\)](#) e é dado ao seguir:

$$\widehat{EBIC} = Dbar + v \log n. \quad (3.11)$$

Onde v é o número de parâmetros e n o número de observações.

Para o cálculo do DIC, EAIC e EBIC códigos na linguagem R foram desenvolvidos.

WAIC

O Watanabe-Akaike Information Criterion (WAIC) foi proposto no trabalho de [Watanabe \(2010\)](#) e é definido ao seguir:

$$\widehat{WAIC} = -2 \left(\widehat{lppd} - \widehat{p}_{WAIC} \right), \quad (3.12)$$

onde \widehat{lppd} é o logaritmo da densidade preditiva pontual que pode ser aproximado para cálculos computacionais da seguinte forma:

$$\widehat{lppd} = \sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p \left(y_i \mid \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} \right) \right). \quad (3.13)$$

para $s = 1, \dots, S$ onde S é o tamanho da amostra a posteriori e $p \left(y_i \mid \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} \right)$ é a densidade de y avaliada no s -ésimo valor de $\boldsymbol{\beta}^{(s)}$ e $\boldsymbol{\delta}^{(s)}$.

Enquanto que:

$$\widehat{p_{WAIC}} = 2 \sum_{i=1}^n \left(\log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)}) \right) - \frac{1}{S} \sum_{s=1}^S \log p(y_i | \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)}) \right). \quad (3.14)$$

PSIS-LOO

O leave-one-out cross-validation (LOO) bayesiano pode ser obtido com os valores de $\boldsymbol{\beta}^{(s)}$ e $\boldsymbol{\delta}^{(s)}$ da distribuição válida, utilizando os índices de importância dados por:

$$r_i^s = \frac{1}{p(y_i | \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)})} \propto \frac{p(\boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} | y_{-i})}{p(\boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} | y)}.$$

Porém a distribuição a posteriori $p(\boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} | y)$ tem menor variância e caudas mais finas do que a distribuição leave-one-out $p(\boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)} | y_{-i})$, tendo falta estabilidade devido a que os índices de importância podem ter variância grande ou infinita. Posteriormente, foi proposto usar o ajuste por máxima verosimilhança da distribuição de Pareto generalizada à cauda superior da distribuição dos índices de importância, e construir um teste para saber se a variância dos índices de importância é finita (KOOPMAN; SHEPHARD; CREAL, 2009).

Por sua vez, a modificação do índice de importância pelo peso truncado proposto por Ionides (2008), garante variância finita e tem a seguinte forma:

$$w_i^s = \min(r_i^s, \sqrt{S\bar{r}}).$$

No trabalho de Vehtari, Gelman e Gabry (2015) mostra-se que a estimação do LOO pode ser melhorada usando Pareto smoothed importance sampling (PSIS), que utiliza um processo de suavização nos índices de importância. O WAIC é assintoticamente igual ao LOO, porém no trabalho de Vehtari, Gelman e Gabry (2016) é demonstrado que o PSIS-LOO apresenta a vantagem de ser mais robusto no caso finito com prioris não informativas ou observações influentes. A forma do PSIS-LOO é dada a seguir:

$$\widehat{PSIS-LOO} = -2 \left(\sum_{i=1}^n \log \left(\frac{\sum_{s=1}^S w_i^s p(y_i | \boldsymbol{\beta}^{(s)}, \boldsymbol{\delta}^{(s)})}{\sum_{s=1}^S w_i^s} \right) \right). \quad (3.15)$$

Analisando o parâmetro estimado \hat{k} da distribuição de Pareto generalizada ajustada pode-se avaliar a confiabilidade da estimação:

- Se $k < \frac{1}{2}$, a variância dos índices de importância é finita. Convergência rápida.

- Se $\frac{1}{2} < k < 1$, a variância dos índices de importância é infinita mas a média dos índices de importância existe. Convergência devagar.
- Se $k > 1$, nem a variância nem a média dos índices de importância existe.

Se $k > 0,7$ recomenda-se o uso de outras medidas ou um modelo mais robusto.

Para o cálculo do WAIC e do PSIS-LOO foram usadas as funções implementadas no pacote loo (VEHTARI; GELMAN; GABRY, 2016).

É importante sinalizar que um modelo é considerado adequado se quando apresentar menores valores dessas medidas.

As medidas DIC, EAIC, e EBIC foram obtidas por códigos próprios na linguagem de programação R; as medidas WAIC e PSIS-LOO foram obtidas das saídas fornecidas pelo pacote loo (VEHTARI; GELMAN; GABRY, 2016) do programa R.

Avaliação preditiva

Uma abordagem para a comparação de modelos é a avaliação preditiva sob o modelo analisado (POWERS, 2011).

Nos modelos de regressão binária propostos é possível definir uma matriz de dupla entrada 2×2 chamada matriz de confusão mostrada na Tabela 8.

Tabela 8 – Matriz de Confusão.

		y_p	
		sucesso	falha
y	sucesso	VP	FN
	falha	FP	VN

Na tabela apresentada y corresponde as respostas binárias na amostra (se $y = 1$ é sucesso e se $y = 0$ é falha), y_p (ou variável predita) corresponde as respostas classificadas pelo modelo sob análise, isto é, se $y_p = 1$ a observação é classificada como sucesso e se $y_p = 0$ a observação é classificada como falha.

A tabela apresenta a distribuição obtida para os elementos da amostra.

Os valores das células são definidos a seguir:

- Verdadeiro positivo (VP): Se a observação é classificada como sucesso e de fato é sucesso.
- Falso negativo (FN): Se a observação é classificada como falha e de fato é sucesso.
- Falso positivo (FP): Se a observação é classificada como sucesso e de fato é falha.

- Verdadeiro negativo (VN): Se a observação é classificada como falha e de fato é falha.

Também $VP + FN + FP + VN = n$, onde n é o tamanho da amostra.

Para obter a matriz de confusão na aplicação foram desenvolvidos códigos no R. Também, a matriz pode ser obtida com a função `confusionMatrix` do pacote `caret` do R (KUHN, 2008).

A partir das classificações detalhadas acima podem-se obter as seguintes medidas de avaliação preditiva (POWERS, 2011):

- Sensibilidade ou razão de verdadeiro positivo (RVP):

$$RVP = \frac{VP}{VP + FN}$$

- Razão de falso negativo (RFN):

$$RFN = \frac{FN}{VP + FN}$$

- Razão de falso positivo (RFP):

$$RFP = \frac{FP}{FP + VN}$$

- Especificidade ou razão de verdadeiro negativo (RVN):

$$RVN = \frac{VN}{FP + VN}$$

A sensibilidade mede a proporção de sucessos que foram corretamente classificados como sucessos. A razão de falso negativo mede a proporção de sucessos que foram erroneamente classificados como falhas. A razão de falso positivo mede a proporção de falhas que tem sido erroneamente classificados como sucessos. A especificidade mede a proporção de falhas que tem sido corretamente classificados como falhas.

Adicionalmente é possível usar a curva característica de operação do receptor (ROC) Hanley e McNeil (1982) a qual é um gráfico que apresenta o comportamento de um classificador binário em diferentes cenários do valor limite para a classificação.

A curva ROC é representada em duas dimensões (X, Y), no eixo Y coloca-se o RVP ou sensibilidade e no eixo X coloca-se o RFP ou 1-especificidade. Cada par de pontos (x, y) vai ser obtido das medidas RVP e RFP segundo as classificações feitas por algum valor limite específico. Assim, precisam-se considerar vários valores limite. Observar que quanto maior for o RVP e menor o RFP, melhor será a classificação feita. Do mesmo jeito, quanto maior for a especificidade ou RVN e menor o RFN, melhor será a classificação.

Para comparar o desempenho dos classificadores pode-se fazer uso da área abaixo da curva ROC ou AUC, *area under the curve*. Hanley e McNeil (1982) sinalizaram que o AUC

tem a propriedade importante de ser equivalente ao teste de Wilcoxon. Ademais, um modelo de classificação binária é considerado adequado se o valor do AUC é próximo a 1, se não for adequado será próximo a 0,5 ([POWERS, 2011](#)).

No trabalho a função `roc.area` do pacote `verification` ([LABORATORY, 2015](#)) do R foi utilizada para o cálculo do AUC e a obtenção dos gráficos.

ESTUDO DE SIMULAÇÃO

Neste Capítulo foi desenvolvido um estudo de simulação para comparar dois algoritmos, Metropolis-Hastings dentro de Gibbs e No-U-Turn Sampler, na estimação bayesiana do modelo de regressão binária com função de ligação potência probit. A comparação foi realizada usando diferentes medidas que avaliem a acurácia e a eficiência na estimação de parâmetros, o tempo computacional e a eficiência por tempo computacional.

4.1 Considerações no estudo de simulação

Objetivo

O objetivo do estudo foi comparar os algoritmos Metropolis-Hastings dentro de Gibbs (MHG) e No-U-Turn Sampler (NUTS). A comparação considerou avaliar a acurácia, eficiência, tempo computacional e eficiência por tempo computacional na amostra válida de dados simulados do modelo de regressão binária considerando função de ligação potência probit. A escolha do modelo foi arbitrária.

Cenários simulados de regressão binária simples com assimetria

Para o estudo de simulação foram considerados dois critérios para definir diferentes cenários simulados de dados de regressão binária com uma covariável, com assimetria. Esses critérios foram o tamanho de amostra ($n = 50; 200; 500; 1000; 2000$) e os valores do parâmetro de assimetria ($\lambda = 1/3; 1; 3$). Dessa forma, em total foram 15 cenários simulados. Seguindo o trabalho de [Bazán, Romeo e Rodrigues \(2014\)](#), em todos os casos os coeficientes de regressão foram fixados considerando os valores $\beta_0 = 0$ e $\beta_1 = 1$ e a covariável foi simulada de uma distribuição Uniforme(-3,3), considerando o conjunto de valores correspondente ao tamanho de amostra de cada cenário. Com essas especificações os valores da variável resposta foram finalmente simulados de uma distribuição de Bernoulli, considerando 50 réplicas.

Especificações de simulação de MCMC

Segundo [Kramer, Calderhead e Radde \(2014\)](#), o HMC (seção 3.1.3) fornece amostras de baixa autocorrelação. Sabendo disso, [Almond \(2014\)](#) sinalizou que menores cadeias podem ser geradas. Levando isso em consideração, decidimos simular sob duas especificações de simulação de MCMC.

Caso 1. Ou especificações de simulação de uma cadeia usual, considerada comumente na literatura. As especificações foram a geração de uma amostra de tamanho 20000, eliminando os primeiros 10000 valores no período do *burnin*; considerando o espaçamento de 10 valores para evitar os efeitos da autocorrelação. O tamanho final de amostra válida sob essas especificações foi 1000.

Caso 2. Ou especificações de simulação de uma cadeia menor, seguindo a sugestão de [Almond \(2014\)](#). As especificações foram a geração de uma amostra de tamanho 4000, retirando no período do *burnin* 2000 valores; sem considerar espaçamento. O tamanho final de amostra válida foi 2000 sob essas condições.

A estimação foi feita considerando os algoritmos NUTS e MHG, implementados nos pacotes Stan ([CARPENTER et al., 2016](#)) e JAGS ([PLUMMER, 2003](#)), respectivamente. Para ambos os algoritmos foram escritos os códigos na linguagem correspondente, apresentados no Anexo C.

Medidas de comparação

Para a comparação dos métodos de estimação foram utilizadas as medidas dadas a seguir segundo os objetivos do estudo de simulação. O método de estimação 1 é o NUTS e o método de estimação 2 é o MHG.

1. Acurácia

A qual é medida usando:

a) EQM

O erro quadrático médio de $\hat{\theta}$ é:

$$\widehat{\text{EQM}} = \frac{1}{M} \sum_{m=1}^M \left[\hat{\theta}^{(m)} - \theta \right]^2,$$

onde M é o número de réplicas na simulação e $\hat{\theta}^{(m)} = \frac{1}{S} \sum_{s=1}^S \theta^{(s)}$ é a média a posteriori do parâmetro θ na amostra válida de S iterações na réplica m . Se o EQM apresenta um valor baixo, existirá boa acurácia, isto é, boa recuperação e precisão de parâmetros ([WALTHER; MOORE, 2005](#)).

b) Razão de EQM

Com essa medida serão comparados os EQM dos dois métodos de estimação do mesmo modelo. A razão de EQM (REQM) entre os métodos 1 e 2 pode ser escrita como:

$$\widehat{\text{REQM}}_{\frac{1}{2}} = \frac{\widehat{\text{EQM}}_1}{\widehat{\text{EQM}}_2}.$$

Se $\widehat{\text{REQM}}_{\frac{1}{2}} < 1$ o método de estimação 1 apresenta melhor acurácia.

2. Eficiência

A qual é medida usando:

a) TEA

Segundo [Kass et al. \(1998\)](#) o tamanho efetivo de amostra (TEA) é definido como:

$$\text{TEA} = \frac{S}{1 + 2 \sum_{r=1}^{\infty} \rho_r},$$

onde $\sum_{r=1}^{\infty} \rho_r$ é a soma de todas as autocorrelações de defasagem e S é o número de iterações da cadeia na simulação. Essa medida representa o tamanho de amostra independente gerada diretamente da distribuição objetivo. A medida foi obtida com a função `ess` da implementação realizada no pacote `batchmeans` ([HARAN; HUGHES, 2016](#)) do R. Quanto maior for o TEA, melhor será o resultado.

No estudo de simulação, o TEA é dado por:

$$\widehat{\text{TEA}} = \frac{1}{M} \sum_{m=1}^M \text{TEA}_m,$$

onde M é o número de réplicas.

b) Razão de TEA

Com essa medida serão comparados os TEA dos dois métodos de estimação do mesmo modelo. A razão de TEA (RTEA) entre os métodos 1 e 2 pode ser escrita como:

$$\widehat{\text{RTEA}}_{\frac{1}{2}} = \frac{\widehat{\text{TEA}}_1}{\widehat{\text{TEA}}_2}.$$

Se $\widehat{\text{RTEA}}_{\frac{1}{2}} > 1$ o método de estimação 1 é mais eficiente.

3. Tempo computacional

O qual é medida usando:

a) Tempo

O tempo computacional dado em segundos. Foi obtido utilizando a função `system.time` do programa R, extraindo o objeto `elapsed`.

b) Razão de tempo

Esta medida compara os tempos dos dois métodos de estimação do mesmo modelo. A razão de tempo (RT) entre os métodos 1 e 2 pode ser escrita como:

$$\widehat{RT}_{\frac{1}{2}} = \frac{\widehat{\text{Tempo}}_1}{\widehat{\text{Tempo}}_2}.$$

Se $\widehat{RT}_{\frac{1}{2}} < 1$ o método de estimação 1 apresenta menor tempo computacional.

4. Eficiência por tempo computacional

A qual é medida usando:

a) TEA/s

Considerando a definição de TEA podemos obter o tamanho efetivo de amostra por segundo.

$$\widehat{\text{TEA/s}} = \frac{\widehat{\text{TEA}}}{\text{seg}}$$

b) Razão de TEA/s

[Sahu \(2002\)](#) utilizou o TEA/s para comparar dois métodos de estimação do mesmo modelo. A razão de eficiência (RE) entre os métodos 1 e 2 é dada a seguir:

$$\widehat{RE}_{\frac{1}{2}} = \frac{(\widehat{\text{TEA/s}})_1}{(\widehat{\text{TEA/s}})_2}.$$

Se $\widehat{RE}_{\frac{1}{2}} > 1$ o método de estimação 1 é mais eficiente por tempo computacional.

Os resultados do estudo de simulação segundoss os valores de REQM, RTEA, RT e RE; e sua análise são apresentados na próxima seção.

Observação. Na análise não está sendo considerada uma avaliação da convergência de ambos os métodos de estimação.

4.2 Resultados

Quando consideramos todos os cenários, com $n = 50$ (Tabela 9), foi observado que:

- A REQM foi maior que 1 na maior parte das vezes, em ambos os casos de simulação. Ou seja, o MHG apresentou acurácia superior comparado ao NUTS.
- A RTEA foi menor que 1 em ambos os casos de simulação. Isso quer dizer que o algoritmo MHG foi melhor do que o algoritmo NUTS ao apresentar maior eficiência. Além do mais, no caso 1 o TEA do NUTS foi entre 32-68% o TEA do MHG, enquanto que, no caso 2, o TEA do NUTS foi entre 9-19% o TEA do MHG.
- A RT foi maior que 1 em ambos os casos de simulação. Ou seja, o algoritmo MHG apresentou menores tempos computacionais do que o algoritmo NUTS. Também vemos que no caso 1, o tempo do NUTS foi 10 vezes o tempo do MHG, enquanto que, no caso 2, o tempo do NUTS foi 45 vezes o tempo do MHG.
- A RE foi menor que 1 em ambos os casos de simulação. Isso quer dizer que o MHG foi mais eficiente por tempo computacional do que o NUTS. Além disso, a eficiência por tempo computacional do MHG com relação ao NUTS foi mais evidente no caso 2 uma vez que a RE tem valores menores do que no caso 1.

Tabela 9 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 50$ nos casos de simulação 1 e 2.

Cenário	Parâmetros	REQM _{NUTS/MHG}		RTEA _{NUTS/MHG}		RT _{NUTS/MHG}		RE _{NUTS/MHG}	
		Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2
1	$\beta_0 = 0$	1,01	0,99	0,55	0,15			0,05	0,003
	$\beta_1 = 1$	1,01	1,24	0,48	0,15	11,37	46,87	0,05	0,003
	$\lambda = 1/3$	1,43	0,67	0,68	0,19			0,06	0,004
2	$\beta_0 = 0$	1,00	1,05	0,49	0,13			0,05	0,003
	$\beta_1 = 1$	1,07	1,09	0,44	0,13	9,81	44,08	0,05	0,003
	$\lambda = 1$	1,01	1,07	0,67	0,18			0,07	0,004
3	$\beta_0 = 0$	13,40	1,38	0,35	0,10			0,04	0,002
	$\beta_1 = 1$	26,73	1,54	0,32	0,09	8,57	43,93	0,03	0,002
	$\lambda = 3$	1,03	1,07	0,63	0,18			0,07	0,004

Quando consideramos todos os cenários, com $n = 200$ (Tabela 10), foi observado que:

- A REQM foi maior que 1 na maior parte das vezes no caso 1, ou seja, o MHG foi superior em acurácia. Por sua vez, a REQM foi menor que 1 na maior parte das vezes no caso 2, ou seja, o NUTS foi superior em acurácia.
- A RTEA foi menor que 1 em ambos os casos de simulação. Então, o MHG foi melhor do que o NUTS ao apresentar maior eficiência Além do que no caso 1 o TEA do NUTS foi entre 47-66% o TEA do MHG, enquanto que, no caso 2 o TEA do NUTS foi entre 12-19% o TEA do MHG.
- A RT foi maior que 1 em ambos os casos de simulação, conseqüentemente o MHG teve menores tempos computacionais do que o NUTS. Além do mais, no caso 1 o tempo do NUTS foi 4,5 vezes o tempo do MHG, enquanto que, no caso 2, o tempo do NUTS foi 13 vezes o tempo do MHG.
- A RE foi menor que 1 em ambos os casos de simulação, conseqüentemente o MHG foi mais eficiente por tempo computacional do que o NUTS. Além disso, a eficiência por tempo computacional do MHG com relação ao NUTS foi mais evidente no caso 2 uma vez que a RE tem valores menores do que no caso 1.

Tabela 10 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 200$ nos casos de simulação 1 e 2.

Cenário	Parâmetros	REQM $_{\frac{NUTS}{MHG}}$		RTEA $_{\frac{NUTS}{MHG}}$		RT $_{\frac{NUTS}{MHG}}$		RE $_{\frac{NUTS}{MHG}}$	
		Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2
4	$\beta_0 = 0$	0,96	0,50	0,55	0,14			0,12	0,011
	$\beta_1 = 1$	1,01	0,89	0,52	0,14	4,37	13,30	0,12	0,010
	$\lambda = 1/3$	1,07	0,20	0,66	0,18			0,15	0,014
5	$\beta_0 = 0$	1,01	0,95	0,53	0,14			0,10	0,011
	$\beta_1 = 1$	1,01	0,93	0,50	0,14	5,26	13,00	0,10	0,011
	$\lambda = 1$	1,00	1,14	0,65	0,19			0,12	0,014
6	$\beta_0 = 0$	1,08	0,94	0,49	0,12			0,11	0,009
	$\beta_1 = 1$	1,09	1,01	0,47	0,12	4,32	13,13	0,11	0,009
	$\lambda = 3$	1,03	1,04	0,63	0,17			0,15	0,013

Quando consideramos todos os cenários, com $n = 500$ (Tabela 11), foi observado que:

- A REQM foi maior que 1 na maior parte das vezes no caso 1, ou seja, o MHG foi superior em acurácia. Por sua vez, a REQM foi menor que 1 na maior parte das vezes no caso 2, ou seja, o NUTS foi superior em acurácia.
- A RTEA foi menor que 1 em ambos os casos de simulação. Dessa forma, o MHG foi melhor do que o NUTS ao apresentar maior eficiência. Além do mais, no caso 1 o TEA do NUTS foi entre 53-67% o TEA do MHG, enquanto que, no caso 2 o TEA do NUTS foi entre 14-19% o TEA do MHG.
- A RT foi maior que 1 em ambos os casos de simulação, conseqüentemente o MHG teve menores tempos computacionais do que o NUTS. Além do que no caso 1 o tempo do NUTS foi 3,7 vezes o tempo do MHG, enquanto que, no caso 2, o tempo do NUTS foi 7,5 vezes o tempo do MHG.
- A RE foi menor que 1 em ambos os casos de simulação, assim, o MHG foi mais eficiente por tempo computacional do que o NUTS. Além disso, a eficiência por tempo computacional do MHG com relação ao NUTS foi mais evidente no caso 2 uma vez que a RE tem valores menores do que no caso 1.

Tabela 11 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 500$ nos casos de simulação 1 e 2.

Cenário	Parâmetros	$REQM_{\frac{NUTS}{MHG}}$		$RTEA_{\frac{NUTS}{MHG}}$		$RT_{\frac{NUTS}{MHG}}$		$RE_{\frac{NUTS}{MHG}}$	
		Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2
7	$\beta_0 = 0$	1,02	0,70	0,60	0,14			0,18	0,02
	$\beta_1 = 1$	1,02	0,62	0,58	0,14	3,38	7,73	0,17	0,02
	$\lambda = 1/3$	1,09	0,38	0,66	0,17			0,20	0,02
8	$\beta_0 = 0$	1,00	0,70	0,58	0,16			0,16	0,02
	$\beta_1 = 1$	0,98	0,70	0,57	0,16	3,67	7,50	0,16	0,02
	$\lambda = 1$	0,99	0,67	0,67	0,19			0,18	0,03
9	$\beta_0 = 0$	1,02	1,05	0,55	0,15			0,15	0,02
	$\beta_1 = 1$	1,03	1,07	0,53	0,15	3,65	7,24	0,15	0,02
	$\lambda = 3$	1,02	1,19	0,61	0,17			0,17	0,02

Quando consideramos todos os cenários, com $n = 1000$ (Tabela 12), foi observado que:

- A REQM foi maior que 1 na maior parte das vezes no caso 1, ou seja, o MHG foi superior em acurácia. Por sua vez, a REQM foi menor que 1 no caso 2, ou seja, o NUTS apresentou acurácia superior, comparado ao MHG.
- A RTEA foi menor que 1 em ambos os casos de simulação. Assim, o MHG foi melhor do que o NUTS ao apresentar maior eficiência. Além disso, no caso 1 o TEA do NUTS foi entre 56-71% o TEA do MHG, enquanto que, no caso 2 o TEA do NUTS foi entre 15-19% o TEA do MHG.
- A RT foi maior que 1 em ambos os casos de simulação, conseqüentemente o MHG teve menores tempos computacionais do que o NUTS. Além disso, no caso 1 o tempo do NUTS foi 3,4 vezes o tempo do MHG, enquanto que, no caso 2, o tempo do NUTS foi 5,5 vezes o tempo do MHG.
- A RE foi menor que 1 em ambos os casos de simulação, dessa forma, o MHG foi mais eficiente por tempo computacional do que o NUTS. Também vemos que a eficiência por tempo computacional do MHG com relação ao NUTS foi mais evidente no caso 2 uma vez que a RE tem valores menores do que no caso 1.

Tabela 12 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 1000$ nos casos de simulação 1 e 2.

Cenário	Parâmetros	REQM $_{\frac{NUTS}{MHG}}$		RTEA $_{\frac{NUTS}{MHG}}$		RT $_{\frac{NUTS}{MHG}}$		RE $_{\frac{NUTS}{MHG}}$	
		Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2
10	$\beta_0 = 0$	0,70	0,48	0,58	0,16			0,18	0,03
	$\beta_1 = 1$	0,85	0,48	0,56	0,15	3,28	5,29	0,17	0,03
	$\lambda = 1/3$	0,30	0,16	0,66	0,19			0,20	0,04
11	$\beta_0 = 0$	1,02	0,66	0,64	0,17			0,19	0,03
	$\beta_1 = 1$	1,02	0,67	0,63	0,17	3,41	5,75	0,18	0,03
	$\lambda = 1$	0,73	0,49	0,71	0,19			0,21	0,03
12	$\beta_0 = 0$	1,02	0,82	0,59	0,15			0,16	0,03
	$\beta_1 = 1$	1,02	0,83	0,59	0,15	3,58	5,61	0,16	0,03
	$\lambda = 3$	1,09	0,99	0,63	0,17			0,18	0,03

Quando consideramos todos os cenários, com $n = 2000$ (Tabela 13), foi observado que:

- A REQM foi menor que 1 em ambos os casos de simulação. Ou seja, o NUTS superou ao MHG em acurácia.
- A RTEA foi menor que 1 em ambos os casos de simulação. Portanto, o MHG foi melhor do que o NUTS ao apresentar maior eficiência. Além do que no caso 1 o TEA do NUTS foi entre 62-73% o TEA do MHG, enquanto que, no caso 2 o TEA do NUTS foi entre 17-23% o TEA do MHG.
- A RT foi maior que 1 em ambos os casos de simulação, conseqüentemente o MHG teve menores tempos computacionais do que o NUTS. Além disso, no caso 1 o tempo do NUTS foi 3,3 vezes o tempo do MHG, enquanto que, no caso 2, o tempo do NUTS foi 4,4 vezes o tempo do MHG.
- A RE foi menor que 1 em ambos os casos de simulação, então, o MHG foi mais eficiente por tempo computacional do que o NUTS. Também vemos que a eficiência por tempo computacional do MHG com relação ao NUTS foi mais evidente no caso 2 uma vez que a RE tem valores menores do que no caso 1.

Tabela 13 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 2000$ nos casos de simulação 1 e 2.

Cenário	Parâmetros	REQM _{NUTS/MHG}		RTEA _{NUTS/MHG}		RT _{NUTS/MHG}		RE _{NUTS/MHG}	
		Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2	Caso 1	Caso 2
13	$\beta_0 = 0$	0,96	0,46	0,65	0,17			0,21	0,04
	$\beta_1 = 1$	0,99	0,54	0,65	0,17	3,09	4,15	0,21	0,04
	$\lambda = 1/3$	0,80	0,11	0,68	0,20			0,22	0,05
14	$\beta_0 = 0$	0,97	0,69	0,73	0,22			0,22	0,05
	$\beta_1 = 1$	0,99	0,72	0,73	0,22	3,39	4,39	0,21	0,05
	$\lambda = 1$	0,94	0,51	0,73	0,23			0,21	0,05
15	$\beta_0 = 0$	0,84	0,79	0,62	0,18			0,18	0,04
	$\beta_1 = 1$	0,85	0,79	0,63	0,18	3,56	4,56	0,18	0,04
	$\lambda = 3$	0,95	0,94	0,65	0,18			0,18	0,04

Pode ser visto que quando o tamanho de amostra n for cada vez maior:

- No caso 1 de simulação, a acurácia usando os algoritmos NUTS e MHG torna-se semelhante. Por outro lado, no caso 2 de simulação, o NUTS apresentou acurácia superior comparado ao MHG.
- A razão do TEA do NUTS com relação ao TEA do MHG aumentou. No caso 1 o TEA do NUTS começou sendo entre 32-68% do TEA do MHG e chegou a ser até 62-73%. Por sua vez, no caso 2 o TEA do NUTS começou sendo entre 9-19% do TEA do MHG e chegou a ser até 17-23%.
- A razão do tempo do NUTS com relação ao tempo do MHG diminuiu. No caso 1 de simulação, o tempo do NUTS começou sendo 10 vezes o tempo do MHG até se tornar 3,3 vez o tempo do MHG. No caso 2 de simulação, o tempo do NUTS começou sendo 45 vezes o tempo do MHG até se tornar 4,4 vezes o tempo do MHG.
- A maior eficiência por tempo computacional do MHG com relação ao NUTS foi se reduzindo, uma vez que essa medida inclui o TEA e o tempo, e elas foram se tornando melhores para o NUTS.

Pode ser concluído que o NUTS superou ao MHG em acurácia, considerando tanto especificações para uma cadeia usual (caso 1) como especificações para uma cadeia menor (caso 2). Por sua vez, o MHG apresentou maior eficiência, menor tempo computacional e maior eficiência por tempo computacional do que o NUTS, considerando tanto especificações para uma cadeia usual como para uma cadeia menor.

Os valores do viés, desvio padrão, erro quadrático médio, tamanho efetivo de amostra, tempo e convergência das cadeias, encontram-se disponíveis no Anexo C.

APLICAÇÃO

Neste Capítulo, é desenvolvida uma aplicação a um conjunto de dados educacionais. Apresentam-se as características do conjunto de dados, os resultados da análise descritiva inicial e os resultados da análise de regressão binária, considerando os diferentes modelos propostos neste trabalho. Todos os modelos propostos no Capítulo 2 são estimados, assim como é realizada a comparação entre eles utilizando critérios de informação e de avaliação preditiva. O modelo com melhor ajuste aos dados é escolhido e é dada a interpretação dos seus parâmetros.

5.1 O conjunto de dados educacionais

No ano de 2004, o Ministério da Educação do Peru (UMC, 2004) realizou a quarta avaliação nacional do desempenho dos estudantes em língua espanhola e matemática, cujo objetivo foi proporcionar informação sobre o desempenho a nível nacional. Três séries das escolas foram avaliadas. Os dados estão disponibilizados (<<http://umc.minedu.gob.pe/evaluacion-nacional-2004/>>) e correspondem a um estudo por amostragem probabilística representativa da população de estudantes.

Neste trabalho, os dados usados foram da avaliação do sexto ano de educação primária do sistema educativo do Peru (equivalente à sexta série do ensino fundamental no Brasil), com meninos e meninas de 11 a 13 anos. A amostra original foi de 13804 estudantes, porém pela ocorrência de dados faltantes de uma parte dos estudantes, a amostra considerada foi de 13259.

As variáveis independentes que foram consideradas neste estudo foram as três variáveis binárias *zona* (rural, urbana), *gestão da escola* (pública, privada) e *sexo* (homem, mulher); e a variável contínua *desempenho em língua espanhola*. Esta última foi a pontuação que o estudante obteve no correspondente teste na avaliação (a pontuação obtida numa escala apresenta média 306,11 e desvio padrão 80,30). Foram escolhidas essas variáveis porque são usualmente consideradas como fatores associados ao desempenho em matemática no Peru, ver, por exemplo,

o trabalho de Bazán, Espinosa e Farro (2002). A variável dependente é uma variável binária que indicará se o estudante teve nível de desempenho adequado em matemática (codificado como 1 pelo estudo) ou não adequado (codificado como 0 pelo estudo).

A descrição das variáveis consideradas é apresentada na Tabela 14.

Tabela 14 – Variáveis na aplicação aos dados educacionais.

Variável	Notação	Tipo	Descrição
Zona	X_1	categórica	Urbana (0) Rural (1)
Gestão	X_2	categórica	Privada (1) Pública (0)
Sexo	X_3	categórica	Homem (1) Mulher (0)
Desempenho em comunicação	X_4	contínua	Pontuação que o estudante obteve.
Nível de desempenho em matemática	Y	categórica	Adequado (1) Não adequado (0)

A equipe de analistas do Ministério de Educação do Peru, utilizando as respostas da prova que avaliou o desempenho dos estudantes em língua espanhola e matemática, ajustou modelos Rasch de crédito parcial (MASTERS, 1982) para obter pontuações de desempenho. Segundo critérios conceituais e empíricos relativos às competências de aprendizado nos estudantes, e estabelecidos pela equipe de analistas (<http://www2.minedu.gob.pe/umc/admin/images/men anexos/men anexos_125.pdf>), as categorias de desempenho foram definidas. Dessa forma, o desempenho de um estudante poderia ser classificado em um de três níveis (prévio, básico e suficiente). Neste trabalho, juntaram-se os níveis prévio e básico, conformando assim a categoria de desempenho não adequado na variável resposta "nível de desempenho em matemática"; o nível suficiente foi considerado como a categoria de desempenho adequado.

5.2 Análise descritiva

Dos 13259 estudantes, 9,98% deles tiveram desempenho adequado em matemática, enquanto que 90,02% deles tiveram desempenho não adequado. Pode ser visto que existe desbalanceamento nos 0's e 1's da variável resposta, dessa forma, o uso de ligações simétricas poderia não ser apropriado na modelagem do conjunto de dados (CHEN; DEY; SHAO, 1999).

Na Tabela 15, são mostradas as porcentagens das covariáveis sexo, gestão da escola e zona, segundo a variável resposta binária "nível do desempenho em matemática". Pode ser visto que, em cada uma das categorias das covariáveis, a porcentagem de estudantes de nível de desempenho adequado é menor do que a porcentagem de estudantes de nível de desempenho não adequado.

Tabela 15 – Porcentagens das variáveis categóricas

Variáveis	Nível de desempenho em matemática	
	Não adequado (0)	Adequado (1)
Zona		
Rural (1)	97,89%	2,11%
Urbana (0)	87,92%	12,08%
Gestão		
Privada (1)	72,19%	27,81%
Pública (0)	94,47%	5,53%
Sexo		
Homem (1)	88,91%	11,09%
Mulher (0)	91,16%	8,84%

Na Tabela 16, para a covariável "desempenho em língua espanhola", são apresentadas as medidas estatísticas descritivas média, mediana, desvio padrão, mínimo e máximo segundo as categorias da variável resposta.

Tabela 16 – Estatísticas da variável quantitativa "desempenho em língua espanhola".

Covariável	Medidas	Nível de desempenho em matemática	
		Não adequado (0)	Adequado (1)
Pontuação em língua espanhola	Média	295,95	397,77
	Mediana	295,00	393,40
	Desvio padrão	75,42	63,12
	Mínimo	9,50	49,50
	Máximo	555,9	555,9

Pode-se observar que a média e a mediana da variável quantitativa "desempenho em língua espanhola" são maiores para os estudantes com desempenho adequado em matemática por 100 unidades do que para os estudantes de desempenho não adequado em matemática. Os desvios padrão apresentam valores semelhantes em cada uma das categorias do nível de desempenho em matemática.

Na Figura 28 é apresentado o boxplot do desempenho em língua espanhola para os estudantes com desempenho não adequado e adequado na avaliação em matemática.

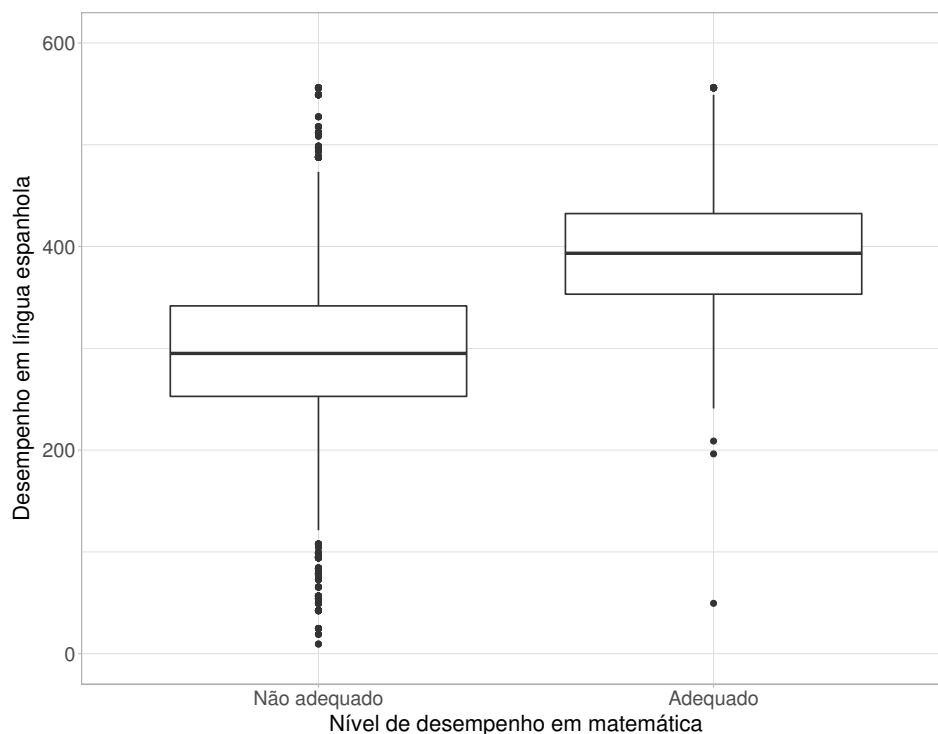


Figura 28 – Boxplot do desempenho em língua espanhola segundo o nível de desempenho em matemática.

Os estudantes com nível de desempenho adequado em matemática têm as pontuações maiores no desempenho em língua espanhola. Também é visto que existem mais valores atípicos no grupo de estudantes de desempenho não adequado em matemática.

5.3 Análise de regressão binária para o desempenho em matemática

Foi visto que todas as covariáveis consideradas na análise descritiva parecem ser importantes para o estudo do desempenho dos estudantes em matemática. Assim, pela análise descrita, as covariáveis parecem estar associadas ao nível de desempenho em matemática. A seguir será detalhada a análise de regressão binária considerando os modelos com funções de ligação potência e reversa de potência.

Processo de estimação

Foram ajustados todos os modelos de regressão binária com ligações potência e reversa de potência apresentados no Capítulo 2 no conjunto de dados educacionais. Para fazer a estimação dos modelos de regressão binária propostos, seguiu-se o procedimento detalhado no Capítulo 3, considerando o algoritmo NUTS implementado no software Stan. Essa escolha é arbitrária e certamente poderia ter sido considerado o algoritmo Metropolis-Hastings dentro de Gibbs no software JAGS, porém, a disponibilidade de algumas funções prontas no Stan para a análise posterior fizeram que fosse escolhido o NUTS. Para o ajuste de cada um dos quinze modelos

(cinco usando ligações comuns, cinco usando ligações potência e cinco usando ligações reversa de potência), obteve-se uma amostra a posteriori de tamanho 8000 e retirando-se os primeiros 4000 valores no período do *burnin*, tendo 4000 valores para os quais considerou-se o espaçamento de tamanho 4 para evitar os efeitos da autocorrelação na amostra, tendo no final uma amostra válida da distribuição a posteriori de 1000 valores para cada caso. Foi verificado em todos os casos, com exceção dos modelos usando funções de ligação PCLL e PCLLR, que as cadeias convergiram considerando o critério do diagnóstico de Geweke (GEWEKE, 1992).

Comparação dos modelos propostos

Depois de finalizar o processo de estimação para cada modelo de regressão binária, obteve-se uma cadeia dos 1000 valores da amostra válida. Considerando essas amostras válidas, obtiveram-se as medidas de informação detalhadas na seção 3.2 para cada caso.

Os resultados são apresentados na Tabela 17, na qual as notações para as ligações correspondem às apresentadas no Capítulo 2:

Tabela 17 – Critérios de informação dos modelos de regressão binária ajustados considerando funções de ligação potência e reversa de potência.

Ligação	DIC	EAIC	EBIC	WAIC	LOO
L	6199,03	6202,23	6232,20	6199,13	6199,12
P	6160,76	6163,76	6193,73	6160,83	6160,82
CLL	6285,13	6288,25	6318,22	6285,59	6285,58
LL	6160,38	6163,26	6193,23	6162,85	*
C	6655,06	6658,23	6688,20	6658,03	6658,03
PL	6142,02	6144,79	6174,76	6143,18	6143,07
PP	6141,94	6154,20	6184,16	6151,74	6151,73
PCLL [†]	249025,4	249033,4	249063,3	249025,7	*
PLL	6160,29	6163,21	6193,18	6162,36	*
PC	8773,80	8780,40	8810,37	8773,37	8773,37
RPL	6137,47	6139,80	6169,77	6137,78	6137,77
RPP	6148,97	6151,80	6181,77	6150,21	6150,17
RPCLL [†]	60664,37	60672,35	60702,32	60664,35	60664,35
RPLL	6285,18	6288,27	6318,24	6285,72	6285,71
RPC	6244,19	6246,35	6276,32	6245,84	6245,83

*: Problemas na estimação do parâmetro k no cálculo do LOO. †: Problemas de convergência.

Segundo os critérios de informação, os modelos mais adequados serão aqueles que tenham menores valores nessas medidas, desse modo escolherão-se os modelos de regressão binária (segundo a ordem do melhor) com funções de ligação RPL, PL, RPP e PP. Note também que o LOO (PSIS-LOO) não foi calculado para os modelos com ligação LL, PCLL e PLL devido a problemas na estimação dessa medida. Além do mais, obtiveram-se valores grandes nos critérios de informação dos modelos de regressão binária com ligação PCLL e RPCLL, os quais serão analisados no futuro em novos estudos de simulação.

Na Tabela 18 pode-se ver a ordem dos quatro modelos com menores medidas de critérios

de informação, segundo qual for o melhor modelo para cada critério. Na ordem, 1 significa que o modelo teve o primeiro melhor valor para esse critério de informação, 2 o segundo melhor valor, e assim sucessivamente. O modelo que obteve melhores valores nos cinco critérios considerados foi o RPL.

Tabela 18 – Ordem na escolha do melhor modelo segundo os critérios de informação.

	DIC	EAIC	EBIC	WAIC	LOO
PL	3	2	2	2	2
RPL	1	1	1	1	1
PP	2	4	4	4	4
RPP	4	3	3	3	3

Avaliação preditiva dos modelos

Para cada um dos quatro modelos escolhidos previamente obtiveram-se as frequências das classificações segundo a matriz de confusão: VP, FN, FP e VN, ilustradas nas Tabelas 32 a 35 do Anexo D. Também obtiveram-se as medidas de avaliação preditiva baseadas nas frequências das classificações das matrizes de confusão: RVP, RFN, RFP e RVN, cujo cálculo foi detalhado na seção 3.2, considerando-se diferentes pontos de corte para a classificação. Isso é ilustrado nas Tabelas 36 a 39 do Anexo D.

Lembre que o RVP é a sensibilidade e o RVN é a especificidade para cada caso. Considerando essas medidas e um maior número de pontos de corte é possível elaborar a curva ROC e obter o AUC (ou o valor da área embaixo da curva ROC), que permite fazer uma escolha do melhor modelo de regressão binária segundo o desempenho que o modelo teve como classificador binário. Para obter as curvas ROC e as medidas do AUC apresentadas a seguir utilizou-se o pacote *verification* (LABORATORY, 2015) do programa R, como foi mencionado na seção 3.2.2. Os valores do AUC dos 4 modelos considerados são apresentados na Tabela 19.

Tabela 19 – AUC dos modelos considerados.

Modelo	AUC
PL	0.8726381
PLR	0.8726878
PP	0.8725706
PPR	0.872634

Observa-se que os valores do AUC para os quatro modelos são próximos a 0,87, esse valor é mais perto de 1 do que de 0,5 e isso indica que todos esses modelos de regressão binária são igualmente adequados. Note-se nesse caso que esse critério não foi adequado para escolher significativamente um modelo entre o total deles.

Nas Figuras 29 a 32 são apresentadas as curvas ROC, as quais mostram-se semelhantes e conseqüentemente nenhum modelo pode ser preferido usando esse critério. As curvas foram obtidas usando o pacote *verification* do R.

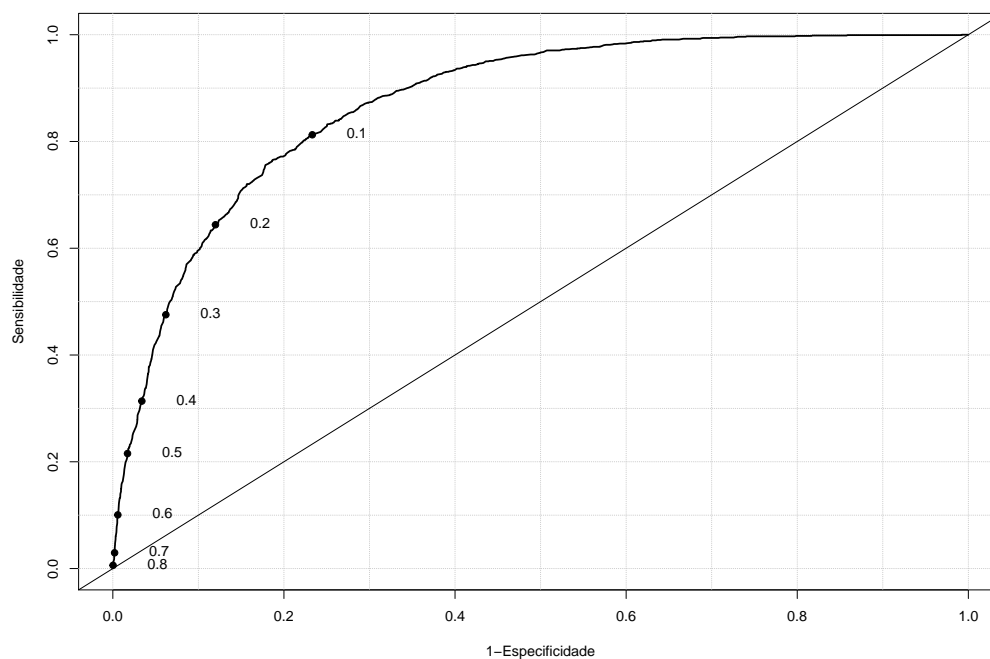


Figura 29 – Curva ROC do classificador binário considerando o modelo de regressão binária com função de ligação PL, $AUC=0,87264$.

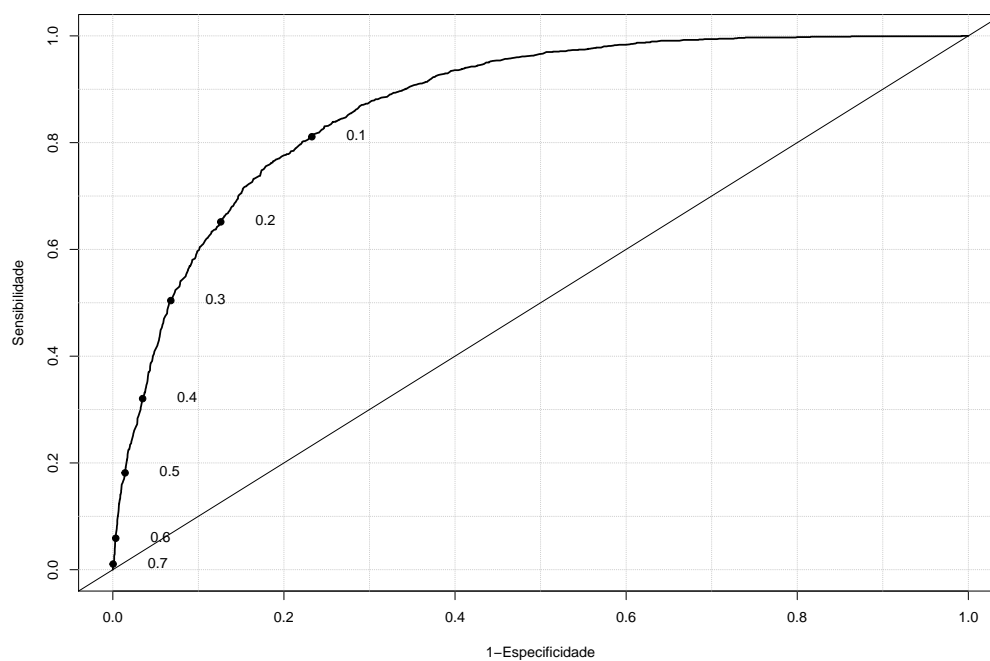


Figura 30 – Curva ROC do classificador binário considerando o modelo de regressão binária com função de ligação RPL, $AUC=0,87269$.

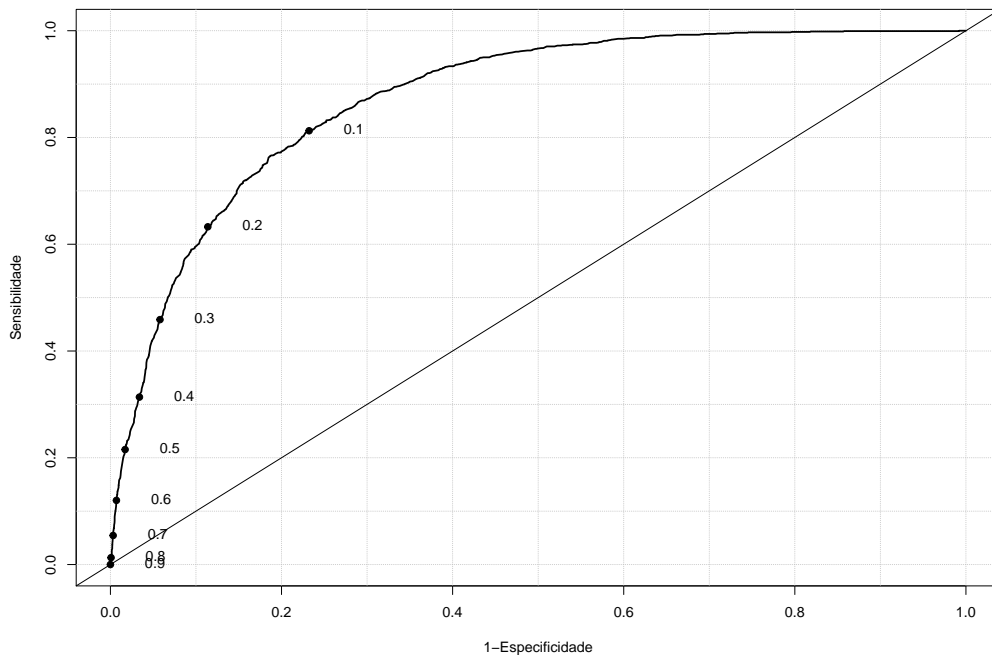


Figura 31 – Curva ROC do classificador binário considerando o modelo de regressão binária com função de ligação PP, AUC=0,87257.

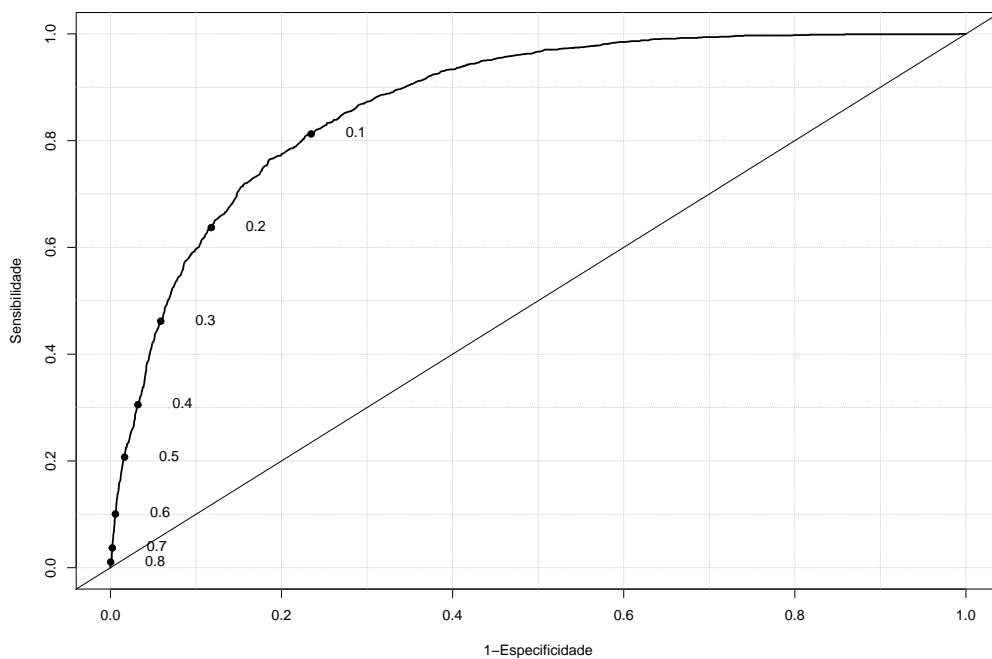


Figura 32 – Curva ROC do classificador binário considerando o modelo de regressão binária com função de ligação RPP, AUC=0,87263.

Modelo final

Levando em consideração os resultados dos critérios de comparação de modelos, o modelo de regressão binária com função de ligação reversa de potência logit foi escolhido. O modelo de regressão binária com função de ligação potência logit foi o segundo melhor, uma breve análise pode ser encontrada no Anexo D.

Na Tabela 20 são apresentados os valores resumo a posteriori dos parâmetros do modelo. Observa-se que todos os parâmetros são significativos pelo fato de que nenhum dos intervalos de credibilidade de 95% inclui o valor zero.

Tabela 20 – Resumo a posteriori dos parâmetros do modelo com função de ligação RPL.

Variáveis	Parâmetros	média	DP	mediana	IC _{95%}	
Intercepto	β_0	-10,651	0,524	-10,626	-11,748	-9,696
Zona	β_1	-0,746	0,189	-0,742	-1,121	-0,378
Gestão	β_2	1,511	0,135	1,504	1,272	1,785
Sexo	β_3	0,661	0,112	0,662	0,459	0,877
Língua espanhola	β_4	0,028	0,002	0,028	0,025	0,032
Assimetria	δ	-1,687	0,148	-1,693	-1,959	-1,369
	λ	0,187	0,029	0,184	0,141	0,254

DP: Desvio padrão. IC: Intervalo de credibilidade.

Finalmente, considerando a análise para explicar o desempenho em matemática e considerando o modelo escolhido e as covariáveis, tem-se o seguinte modelo final:

$$Y_i \sim \text{Bernoulli}(\hat{p}_i); i = 1, \dots, 13259, \quad (5.1)$$

com:

$$\hat{p}_i = 1 - \left(\frac{e^{-(-10.651 - 0.746X_1 + 1.511X_2 + 0.661X_3 + 0.028X_4)}}{1 + e^{-(-10.651 - 0.746X_1 + 1.511X_2 + 0.661X_3 + 0.028X_4)}} \right)^{0.187}. \quad (5.2)$$

Da expressão das probabilidades estimadas considerando a equação 5.2, observa-se o seguinte:

- O parâmetro associado à variável zona (X_1) é negativo, isto quer dizer que os estudantes de zona rural têm menor probabilidade de ter nível de desempenho adequado em matemática do que os estudantes de zona urbana.
- O parâmetro associado à variável gestão (X_2) é positivo, isto quer dizer que os estudantes de escola privada têm maior probabilidade de ter nível de desempenho adequado em matemática do que os estudantes de escola pública.
- Por sua vez, o parâmetro associado à variável sexo (X_3) é positivo, isto quer dizer que, de acordo com este modelo, os homens têm maior probabilidade de ter nível de desempenho adequado em matemática do que as mulheres.

- Da mesma maneira, o parâmetro associado a desempenho em língua espanhola (X_4) é positivo e indica que quando o estudante tiver maior pontuação nessa variável, maior será a probabilidade de ele ter nível de desempenho adequado em matemática.
- O valor do parâmetro de forma está contido em $0 < \lambda < 1$. Segundo o que foi visto no Capítulo 2, nesse caso $\lambda = 0,187$ é um valor coerente à proporção de 1's na variável resposta (menor que 0,5). Além disso, na assimetria o modelo RPL terá A_O positiva.

A interpretação anterior também pode ser analisada na Tabela 21 que inclui o valor das probabilidades estimadas segundo os possíveis valores das variáveis categóricas e níveis de desempenho em língua espanhola. Esses níveis são as médias dos valores da variável desempenho em língua espanhola categorizada segundo critérios do Ministério de Educação do Peru. Dessa forma, pode-se saber o valor das probabilidades de ter nível de desempenho adequado em matemática segundo os valores de todas as covariáveis.

Tabela 21 – Probabilidade estimada de ter nível de desempenho adequado em matemática segundo perfis de estudantes e desempenho em língua espanhola.

Perfil	Zona X_1	Gestão X_2	Sexo X_3	Desempenho em língua espanhola X_4			
				<Prévio (0)	Prévio (1)	Básico (2)	Suficiente (3)
1	Rural	Pública	Mulher	<0,001	0,005	0,030	0,194
2	Rural	Pública	Homem	0,001	0,010	0,054	0,265
3	Urbana	Pública	Mulher	0,001	0,011	0,058	0,275
4	Rural	Privada	Mulher	0,002	0,023	0,105	0,359
5	Urbana	Pública	Homem	0,002	0,021	0,097	0,384
6	Rural	Privada	Homem	0,004	0,041	0,161	0,429
7	Urbana	Privada	Mulher	0,004	0,044	0,169	0,438
8	Urbana	Privada	Homem	0,008	0,077	0,238	0,501

Na Tabela 21, pode ser visto que a probabilidade de ter nível de desempenho adequado em matemática incrementa conforme o desempenho em língua espanhola incrementa, em todos os perfis. Também é visto que o perfil 1 apresenta menores probabilidades. Esse perfil corresponde a uma estudante ($X_3 = 0$) de escola pública ($X_2 = 0$) em zona rural ($X_1 = 1$). Como foi mencionado anteriormente, estudantes com essas características terão menor probabilidade de ter nível de desempenho adequado em matemática. A Figura 33 também mostra o comportamento dos diferentes perfis de estudantes.

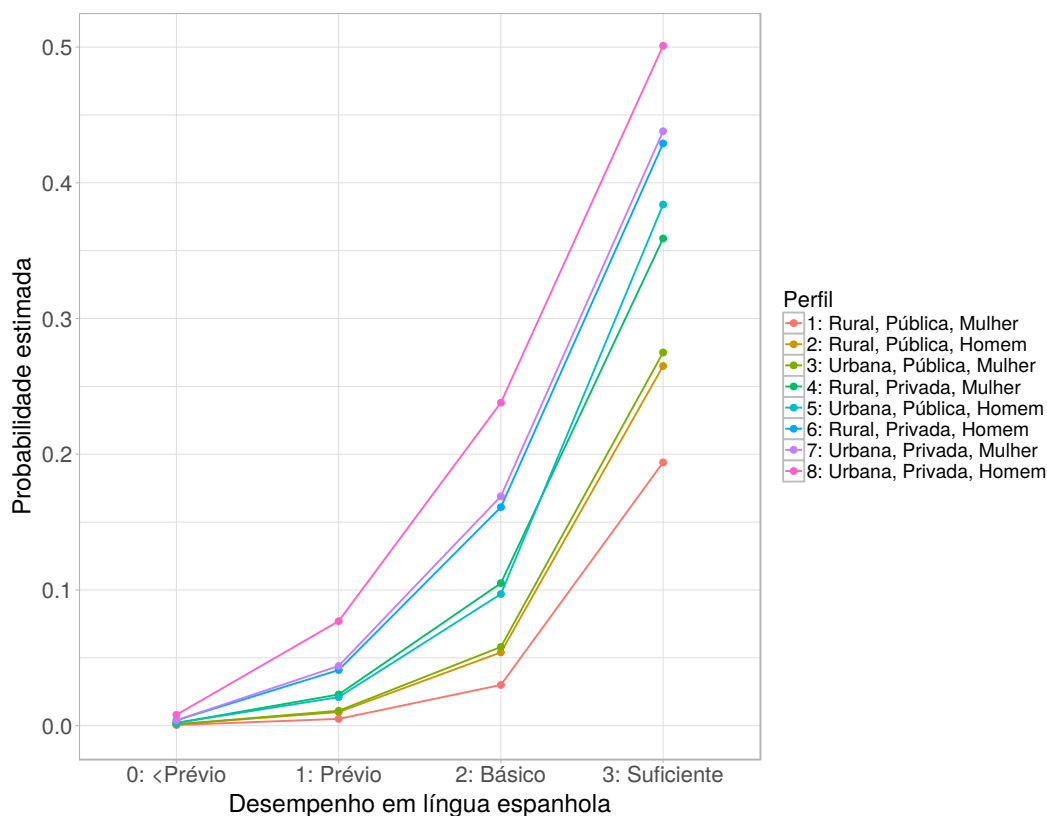


Figura 33 – Perfis de estudantes.

As diferenças no nível de desempenho em matemática segundo o sexo dos estudantes no Peru tanto de ensino fundamental como de ensino médio foi estudada anteriormente em [Bazán, Espinosa e Farro \(2002\)](#). As conclusões nesse trabalho são coerentes com o que foi encontrado nesta aplicação, com resultados favoráveis ao sexo masculino. Para entender esse resultado, várias pesquisas sugerem o rol da socialização do gênero ([ECCLES; JACOBS, 1986](#)), ou o fato de que as mulheres experimentam maior ansiedade na matemática do que os homens ([MEECE; WIGFIELD; ECCLES, 1990](#)). Além disso, estudos recentes como o de [Devine et al. \(2012\)](#) não encontraram diferenças no desempenho relacionado ao sexo apesar da diferença de ansiedade, concluindo que as mulheres poderiam ter melhor desempenho do que os homens se a ansiedade pudesse ser atenuada. Além do mais, no trabalho de [Bieg et al. \(2015\)](#) concluiu-se que devido a estereótipos sobre diferenças no desempenho em matemática segundo o sexo, as estudantes mulheres podem acreditar que estão mais ansiosas do que realmente estão em relação à matemática. Esse trabalho enfatizou a tarefa dos professores e das famílias dos estudantes em reduzir esses estereótipos. Isso significa que eventuais conclusões baseadas nos resultados apresentados neste capítulo deverão levar em conta os trabalhos citados acima.

Além do mais, para o Peru, no trabalho de [Cueto, Jacoby e Pollitt \(2013\)](#), usando outra metodologia, concluiu-se que as diferenças de desempenho entre homens e mulheres foram maiores nas escolas de zona rural, que encontram-se também em desvantagem frente a escolas de zona urbana. Adicionalmente, em [Cox \(2000\)](#), encontraram-se diferenças entre o desempenho

de estudantes segundo a zona rural ou urbana da escola do estudante na Austrália. Por sua vez, as diferenças que existem na atitude à matemática segundo o tipo de gestão de escola na Espanha, foram analisadas em [Mato e Torre \(2009\)](#).

Dado que foram utilizados dados obtidos numa amostragem probabilística, os resultados são generalizáveis para estudantes da sexta série nas escolas do Peru daquele ano.

CONSIDERAÇÕES FINAIS

Nesta dissertação, os modelos de regressão binária foram revisados e estudados, usando ligações assimétricas potência e reversa de potência sob a abordagem bayesiana. Neste tipo de modelos, a estimação geralmente é feita considerando o algoritmo Metropolis-Hastings ou o amostrador de Gibbs. A contribuição do trabalho é considerar tanto o algoritmo Monte Carlo Hamiltoniano na extensão No-U-Turn Sampler como o algoritmo Metropolis-Hastings dentro de Gibbs. Códigos e rotinas foram desenvolvidas nos softwares R, JAGS e Stan (ver Anexo B) que no futuro poderão ser utilizadas para desenvolver um pacote no R.

Estudos de simulação foram desenvolvidos com o objetivo de avaliar acurácia, eficiência, tempo computacional e eficiência por tempo computacional. Conclui-se que na estimação de modelos de regressão binária com funções de ligação potência e reversa de potência, o algoritmo No-U-Turn Sampler superou em acurácia ao algoritmo Metropolis-Hastings dentro de Gibbs na maior parte das vezes. Por sua vez, o algoritmo Metropolis-Hastings dentro de Gibbs superou em eficiência, tempo computacional e eficiência por tempo computacional ao algoritmo No-U-Turn Sampler.

Dados educacionais do Ministério de Educação do Peru foram utilizados para ilustrar a utilidade dos modelos propostos. O ajuste desses modelos teve a finalidade de determinar a relação entre o nível de desempenho dos estudantes em matemática e um conjunto de covariáveis como zona e gestão da escola, sexo do estudante e o seu desempenho em língua espanhola. Os cálculos dos critérios de comparação de modelos foram realizados segundo as medidas de informação e de avaliação preditiva. Foi observado que as medidas de informação são mais efetivas para a comparação de modelos do que a medida de avaliação preditiva baseada na área embaixo da curva ROC (AUC). Na aplicação aos dados educacionais, observou-se que os modelos propostos de regressão binária com funções de ligação potência e reversa de potência obtiveram melhor ajuste do que os modelos de regressão binária usando ligações comuns. O modelo escolhido foi o modelo de regressão binária com função de ligação reversa de potência

logit. Concluiu-se que fatores como gestão e zona da escola, sexo e desempenho em língua espanhola são todos importantes na análise do nível de desempenho em matemática. Estudantes homens, de escolas privadas em zonas urbanas com maior desempenho em língua espanhola terão maior probabilidade de ter nível de desempenho adequado em matemática. Esses resultados são coerentes com os de pesquisas educativas na avaliação do desempenho em matemática. As diferenças no desempenho em matemática segundo o sexo podem ser reflexo da realidade do ano de 2004. Atualmente, essa realidade pode estar mudando. Também pode ser explicado porque as mulheres experimentam maior ansiedade na matemática, possivelmente devido a estereótipos sobre as diferenças no desempenho em matemática segundo o sexo.

Como desdobramento futuro está prevista a disponibilização da biblioteca de software correspondente visando fornecer alternativas às ferramentas de estimação com funções de ligação tradicionais, possibilitando um melhor ajuste para os dados que são modelados pela regressão binária. Além disso, é de interesse desenvolver métodos de estimação clássica, análise de resíduos e outros diagnósticos para os modelos propostos. A extensão para dados mistos e modelos de resposta ordinal também pode ser direcionada.

REFERÊNCIAS

- ABANTO-VALLE, C. A.; BAZÁN, J. L.; SMITH, A. C. **State space mixed models for binary responses with skewed inverse links using JAGS**. Rio de Janeiro, Brazil, 2014. Citado nas páginas 18, 25 e 28.
- AGRESTI, A. **An introduction to categorical data analysis**. Hoboken (N.J.): Wiley-Interscience, 2007. (Wiley series in probability and statistics). ISBN 0-471-22618-1. Citado na página 18.
- ALMOND, R. G. A comparison of two mcmc algorithms for hierarchical mixture models. In: **Proceedings of the Eleventh UAI Conference on Bayesian Modeling Applications Workshop - Volume 1218**. Aachen, Germany, Germany: CEUR-WS.org, 2014. (BMAW'14), p. 1–19. Citado na página 64.
- ARANDA-ORDAZ, F. J. On two families of transformations to additivity for binary response data. **Biometrika**, v. 68, n. 2, p. 357–363, 1981. Citado na página 17.
- BAZÁN, J.; ESPINOSA, G.; FARRO, C. Rendimiento y actitudes hacia la matemática en el sistema escolar peruano. MECEP. Programa Especial Mejoramiento de la Calidad de la Educación Peruana, 2002. Citado nas páginas 74 e 83.
- BAZÁN, J.; TORRES-AVILÉS, F.; SUZUKI, A.; LOUZADA, F. Power and reversal power links for binary regressions: An application for motor insurance policyholders. **Applied Stochastic Models in Business and Industry**, John Wiley & Sons, Ltd, p. n/a–n/a, 2016. Citado nas páginas 18, 23, 25 e 28.
- BAZÁN, J. L.; ROMEO, J. S.; RODRIGUES, J. Bayesian skew-probit regression for binary response data. **Brazilian Journal of Probability and Statistics**, v. 28, n. 4, p. 467–482, 2014. Citado nas páginas 18, 28, 52 e 63.
- BIEG, M.; GOETZ, T.; WOLTER, I.; HALL, N. C. Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. **Frontiers in psychology**, Frontiers, v. 6, p. 1404, 2015. Citado na página 83.
- BROOKS, S. P. **Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde**. Hoboken: Wiley-Blackwell, 2002. Citado na página 58.
- CARPENTER, B.; GELMAN, A.; HOFFMAN, M.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M. A.; GUO, J.; LI, P.; RIDDELL, A. Stan: A probabilistic programming language. **Journal of Statistical Software**, v. 10, 2016. Citado nas páginas 55 e 64.
- CHEN, M.-H.; DEY, D. K.; SHAO, Q.-M. A new skewed link model for dichotomous quantal response data. **Journal of the American Statistical Association**, v. 94, n. 448, p. 1172–1186, 1999. Citado nas páginas 17, 36 e 74.
- COLLETT, D. **Modelling Binary Data**. London, New York: Chapman & Hall, 2003. (Texts in Statistical Science). ISBN 9781584883241. Citado na página 17.

- COX, P. Regional and gender differences in mathematics achievement. **Journal of Research in Rural Education**, v. 16, n. 1, p. 22–29, 2000. Citado na página 83.
- CUETO, S.; JACOBY, E.; POLLITT, E. Rendimiento de niños y niñas de zonas rurales y urbanas del Perú. **Revista de psicología**, v. 15, n. 1, p. 115–133, 2013. Citado na página 83.
- CZADO, C.; SANTNER, T. J. The effect of link misspecification on binary regression inference. **Journal of Statistical Planning and Inference**, v. 33, n. 2, p. 213–231, 1992. Citado na página 17.
- DEVINE, A.; FAWCETT, K.; SZÚCS, D.; DOWKER, A. Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. **Behavioral and brain functions**, BioMed Central, v. 8, n. 1, p. 33, 2012. Citado na página 83.
- DEY, D. K.; GOSH, S. K.; MALLICK, B. K. **Generalized linear models: a Bayesian perspective**. New York: Marcel Dekker Inc, 2000. ISBN 978-0-824-790-349. Citado na página 18.
- DUANE, S.; KENNEDY, A.; PENDLETON, B. J.; ROWETH, D. Hybrid monte carlo. **Physics Letters B**, v. 195, n. 2, p. 216–222, 1987. Citado nas páginas 18 e 55.
- ECCLES, J. S.; JACOBS, J. E. Social forces shape math attitudes and performance. **Signs: Journal of women in culture and society**, University of Chicago Press, v. 11, n. 2, p. 367–380, 1986. Citado na página 83.
- GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for Bayesian models. **Statistics and Computing**, v. 24, n. 6, p. 997–1016, 2014. Citado na página 57.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 6, n. 6, p. 721–741, 1984. Citado nas páginas 18 e 53.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. M.; BERGER, J.; DAWID, A. P.; SMITH, J. F. M. (Ed.). **Bayesian Statistics 4**. Oxford: Oxford University Press, 1992. p. 169–193. Citado nas páginas 77 e 101.
- GILKS, W. R.; BEST, N. G.; TAN, K. K. C. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, [Wiley, Royal Statistical Society], v. 44, n. 4, p. 455–472, 1995. ISSN 00359254, 14679876. Citado na página 53.
- GILKS, W. R.; WILD, P. Adaptive Rejection Sampling for Gibbs Sampling. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, [Wiley, Royal Statistical Society], v. 41, n. 2, p. 337–348, 1992. ISSN 00359254, 14679876. Citado na página 53.
- GUERRERO, V. M.; JOHNSON, R. A. Use of the box-cox transformation with binary response models. **Biometrika**, v. 69, n. 2, p. 309–314, 1982. Citado na página 17.
- GUPTA, R. D.; GUPTA, R. C. Analyzing skewed data by power normal model. **TEST**, v. 17, n. 1, p. 197–210, 2008. Citado nas páginas 17, 28 e 42.
- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. **Radiology**, v. 143, n. 1, p. 29–36, 1982. Citado na página 61.

HARAN, M.; HUGHES, J. **batchmeans: Consistent Batch Means Estimation of Monte Carlo Standard Errors**. Denver, CO, 2016. R package version 1.0-3. Citado na página 65.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 57, n. 1, p. 97–109, 1970. ISSN 00063444. Citado na página 53.

HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. **Journal of Machine Learning Research**, v. 15, p. 1593–1623, 2014. Citado nas páginas 55 e 57.

IONIDES, E. Truncated importance sampling. **Journal of Computational and Graphical Statistics**, American Statistical Association, v. 17, n. 2, p. 295–311, 6 2008. ISSN 1061-8600. Citado na página 59.

JIANG, X.; DEY, D. K.; PRUNIER, R.; WILSON, A. M.; HOLSINGER, K. E. A new class of flexible link functions with application to species co-occurrence in cape floristic region. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 7, n. 4, p. 2180–2204, 2013. ISSN 19326157. Citado na página 52.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. 2nd. ed. Hoboken: Wiley & Sons, 1995. v. 2. ISBN 978-0-471-58494-0. Citado na página 28.

JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado na página 17.

KASS, R. E.; CARLIN, B. P.; GELMAN, A.; NEAL, R. M. Markov chain monte carlo in practice: A roundtable discussion. **The American Statistician**, v. 52, n. 2, p. 93–100, 1998. Citado na página 65.

KIM, S.; CHEN, M.-H.; DEY, D. K. Flexible generalized t-link models for binary response data. **Biometrika**, v. 95, n. 1, p. 93–106, 2008. Citado na página 17.

KOOPMAN, S. J.; SHEPHARD, N.; CREAL, D. Testing the assumptions behind importance sampling. **Journal of Econometrics**, Elsevier, v. 149, n. 1, p. 2–11, 2009. Citado na página 59.

KRAMER, A.; CALDERHEAD, B.; RADDE, N. Hamiltonian monte carlo methods for efficient parameter estimation in steady state dynamical systems. **BMC Bioinformatics**, v. 15, n. 1, p. 253, 2014. ISSN 1471-2105. Citado na página 64.

KUHN, M. Building predictive models in r using the caret package. **Journal of Statistical Software**, v. 28, n. 1, p. 1–26, 2008. ISSN 1548-7660. Citado na página 61.

LABORATORY, N. R. A. **verification: Weather Forecast Verification Utilities**. CO, USA, 2015. R package version 1.42. Disponível em: <<https://CRAN.R-project.org/package=verification>>. Citado nas páginas 62 e 78.

LAN, S. **Advanced Bayesian Computational Methods Through Geometric Techniques**. Tese (Doutorado), Long Beach, CA, USA, 2013. AAI3605182. Citado na página 18.

MASTERS, G. A Rasch model for partial credit scoring. **Psychometrika**, v. 47, p. 149–174, 1982. Citado na página 74.

- MATO, M. D.; TORRE, E. de la. Evaluación de las actitudes hacia las matemáticas y el rendimiento académico. Sociedad Española de Investigación en Educación Matemática, SEIEM, 2009. Citado na página 84.
- MEECE, J. L.; WIGFIELD, A.; ECCLES, J. S. Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. **Journal of educational psychology**, American Psychological Association, v. 82, n. 1, p. 60, 1990. Citado na página 83.
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, 1953. Citado nas páginas 18 e 53.
- MOORS, J. J. A.; WAGEMAKERS, R. T. A.; COENEN, V. M. J.; HEUTS, R. M. J.; JANSSENS, M. J. B. T. Characterizing systems of distributions by quantile measures. **Statistica Neerlandica**, Blackwell Publishing Ltd, v. 50, n. 3, p. 417–430, 1996. Citado na página 42.
- MORGAN, B. J. T. Observations on quantitative analysis. **Biometrics**, v. 39, p. 879–886, 1983. Citado na página 17.
- NAGLER, J. Scobit: An alternative estimator to logit and probit. **American Journal of Political Science**, v. 38, n. 1, p. 230–255, 1994. Citado nas páginas 17 e 28.
- NEAL, R. M. **Probabilistic Inference Using Markov Chain Monte Carlo Methods**. Toronto, Canada, 1993. Citado na página 94.
- NEAL, R. M. **Bayesian learning for neural networks**. New York, Berlin, Paris: Springer, 1996. (Lecture notes in statistics). Revision of the Ph.D. thesis : Bayesian learning for neural networks, Department of Computer Science, University of Toronto, 1995. Citado na página 18.
- NEAL, R. M. MCMC using Hamiltonian dynamics. **ArXiv e-prints**, 2012. Citado nas páginas 55, 56, 57, 93, 94 e 96.
- NTZOUFRAS, I. **Bayesian Modeling Using WinBUGS**. Hoboken: Wiley, 2009. ISBN 978-0-470-14114-4. Citado na página 18.
- PLUMMER, M. **JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling**. 2003. Citado nas páginas 54 e 64.
- POWERS, D. M. W. Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation. **Journal of Machine Learning Technologies**, v. 2, p. 37–63, 2011. Citado nas páginas 60, 61 e 62.
- PRENTICE, R. L. A generalization of the probit and logit methods for dose response curves. **Biometrics**, v. 32, n. 4, p. 761–768, 1976. Citado na página 17.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>. Citado na página 22.
- SAHU, S. K. Bayesian estimation and model choice in item response models. **Journal of Statistical Computation and Simulation**, v. 72, n. 3, p. 217–232, 2002. Citado na página 66.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 64, n. 4, p. 583–639, 2002. Citado na página 57.

STUKEL, T. A. Generalized logistic models. **Journal of the American Statistical Association**, v. 83, p. 426–431, 1988. Citado na página 17.

UMC. **Evaluación Nacional 2004 na responsabilidade da Oficina de Medición de la Calidad de los Aprendizajes do Ministério de Educação do Peru**. 2004. Disponível em: <<http://umc.minedu.gob.pe/evaluacion-nacional-2004/>>. Acesso em: 29/08/2016. Citado na página 73.

VEHTARI, A.; GELMAN, A.; GABRY, J. Pareto smoothed importance sampling. **arXiv pre-print arXiv:1507.02646**, 2015. Citado na página 59.

VEHTARI, A.; GELMAN, A.; GABRY, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. **Statistics and Computing**, 2016. Citado nas páginas 59 e 60.

WALTHER, B. A.; MOORE, J. L. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. **Ecography**, Munksgaard International Publishers, v. 28, n. 6, p. 815–829, 2005. ISSN 1600-0587. Citado na página 64.

WANG, X.; DEY, D. K. Generalized extreme value regression for binary response data: An application to b2b electronic payments system adoption. **Annals of Applied Statistics**, v. 4, n. 4, p. 2000–2023, 2010. Citado na página 17.

WATANABE, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of Machine Learning Research**, v. 11, p. 3571–3594, 2010. Citado na página 58.

REVISÃO DO MÉTODO MONTE CARLO HAMILTONIANO

Neste anexo é detalhado o algoritmo Monte Carlo Hamiltoniano, baseado principalmente no trabalho de [Neal \(2012\)](#).

A.1 Dinâmica Hamiltoniana

Para compreender a dinâmica Hamiltoniana a interpretação física pode fornecer de ideias iniciais. Como apresentado em [Neal \(2012\)](#), consideram-se duas dimensões, assume-se um disco sem atrito que desliza-se sobre uma superfície de altura variável. O sistema é dado pela *posição* do disco (vetor q) e pelo *momento* do disco (vetor p). A *energia potencial*, $U(q)$, do disco é proporcional à altura da superfície na posição atual, e a *energia cinética*, $K(p)$, é igual a $|p|^2 / (2m)$, sendo m a massa do disco. No sistema, o disco desliza-se a velocidade constante igual a p/m . Se o disco estiver subindo a superfície, o momento do disco permite que ele continue subindo, a energia cinética vai diminuir até ser zero e a energia potencial vai aumentar até isso acontecer, nesse momento o disco começa descer e a energia cinética aumenta enquanto a energia potencial diminui.

Equações Hamiltonianas

Considerando o vetor de posição q e o vetor de momento p , no método MCH usa-se a função do Hamiltoniano que é dada por:

$$H(q, p) = U(q) + K(p) = U(q) + p^T M^{-1} p / 2.$$

Onde $U(q)$ é a energia potencial igual a menos o logaritmo da função de densidade de probabilidade da qual vai-se amostrar, q , mais alguma constante. $K(p)$ é a energia cinética, com

M , a matriz de massa $d \times d$ simétrica definida positiva. Os vetores q e p são de dimensão d de forma que o espaço no total tem dimensão $2d$.

As derivadas parciais do Hamiltoniano mostram como q e p mudam no tempo t de acordo as equações Hamiltonianas. Considerando as definição anterior de $H(q, p)$:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} = [M^{-1}p]_i,$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i},$$

para $i = 1, \dots, d$. Determinar a posição e velocidade após a evolução do sistema para um intervalo de tempo s desde um estado no tempo t ao estado $t + s$ pode se conseguir com a solução destas equações Hamiltonianas fixados os valores iniciais $[q(t), p(t)]$ que fornecerão a posição e velocidade em $t + s$ $[q(t + s), p(t + s)]$.

A dinâmica Hamiltoniana satisfaz as propriedades de reversibilidade, conservação, preservação do volume e symplecticidade que são detalhadas em [Neal \(1993\)](#).

Discretização das equações Hamiltonianas: Método Leapfrog

As equações Hamiltonianas devem ser aproximadas pela discretização do tempo usando para isso um passo de tamanho ε e o número de passos L . Dessa forma o MCH poderá ser implementado computacionalmente. O método que da melhores resultados é o leapfrog ([NEAL, 2012](#)). Por simplicidade assume-se que M é diagonal com elementos m_1, \dots, m_d , assim:

$$K(p) = \sum_{i=1}^d \frac{p_i^2}{2m_i}.$$

O método é apresentado ao seguir:

$$p_i(t + \varepsilon/2) = p_i(t) - (\varepsilon/2) \frac{\partial U(q(t))}{\partial q_i},$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{p_i(t + \varepsilon/2)}{m_i},$$

$$p_i(t + \varepsilon) = p_i(t + \varepsilon/2) - (\varepsilon/2) \frac{\partial U(q(t + \varepsilon))}{\partial q_i}.$$

Este método começa com a metade de um passo $\varepsilon/2$ para p , depois fazendo uso dos novos valores de p segue com um passo completo ε para q , para finalizar usa os novos valores de q para seguir com a metade do passo $\varepsilon/2$ para p .

Na programação do MCH fazendo uso do método leapfrog precisa-se especificar os parâmetros ε e L , que são o tamanho do passo e o número de passos, respectivamente. Se

não forem especificados os valores apropriados pode-se chegar a situações nas quais existe gasto computacional ou caminhos aleatórios não efetivos. Para prevenir essas situações esses parâmetros devem ser calibrados.

A.2 MCMC na dinâmica Hamiltoniana

Para simular uma cadeia de Markov da distribuição alvo em cada valor da amostra a posteriori obtêm-se a variável do momento e depois se faz uma atualização de Metropolis com a proposta dada pela dinâmica Hamiltoniana. As variáveis de momento são introduzidas artificialmente, uma para cada variável de posição. As variáveis de posição na função de energia potencial são relacionadas à distribuição da qual quer-se amostrar usando o conceito de *distribuição canônica* da estatística mecânica que é apresentada ao seguir:

$$P(x) = \frac{1}{Z} \exp(-E(x)/T),$$

onde X segue a *distribuição canônica*, $E(x)$ é uma função de energia para um estado x de um sistema físico, T é a temperatura do sistema e Z é a constante de normalização.

O Hamiltoniano é uma função de energia conjunta da posição (q) e o momento (p), então:

$$P(q, p) = \frac{1}{Z} \exp(-H(q, p)/T) = \frac{1}{Z} \exp(-U(q)/T) \exp(-K(p)/T),$$

q e p são independentes e cada uma tem a distribuição canônica.

Na abordagem bayesiana é de interesse obter a distribuição a posteriori, portanto os parâmetros dela terão o lugar das variáveis de posição. A distribuição a posteriori é apresentada como uma distribuição canônica com $T = 1$ usando a energia potencial:

$$U(q) = -\log[\pi(q)L(q|D)],$$

onde $\pi(q)$ é a distribuição a priori, $L(q|D)$ é a função de verossimilhança dos dados D .

Algoritmo MCH

A energia cinética quadrática é usada no MCH de forma que p tem distribuição Gaussiana multivariada de média zero com os componentes de p independentes, onde cada i tem variância m_i , neste caso:

$$\mathbf{p} \sim N_d(\mathbf{0}, \text{diag}(m_1, \dots, m_d)),$$

$$K(p) = \sum_{i=1}^d \frac{p_i^2}{2m_i}.$$

O algoritmo MCH tem dois passos, no primeiro muda o momento, no segundo mudam a posição e o momento. Os dois passos deixam a distribuição canônica invariante.

No primeiro passo, valores novos para as variáveis de momento são gerados da Gaussiana aleatoriamente, e independentemente das variáveis de posição atuais. Na equação de energia cinética as d variáveis de momento são independentes. Nesse passo q não muda.

No segundo passo, usa-se a dinâmica Hamiltoniana para propor um novo estado, começa-se desde o estado atual (q, p) e usa-se o método leapfrog para L passos de tamanho ε . Depois as variáveis de momento obtidas são negadas para obter um estado proposto (q^*, p^*) que é aceito como o novo estado da cadeia de Markov com probabilidade:

$$\min [1, \exp(-H(q^*, p^*) + H(q, p))]$$

Se a proposta é rejeitada o novo estado é o mesmo que o estado atual. Para maiores detalhes do algoritmo MCH revisar [Neal \(2012\)](#).

CÓDIGOS ILUSTRATIVOS DOS MODELOS NO STAN E JAGS

Neste Anexo fornecem-se os códigos nas linguagens do Stan e JAGS utilizados na simulação do modelo de regressão binária com função de ligação potência probit do Capítulo 4.

B.1 Códigos no Stan

O seguinte código corresponde a implementação na linguagem do Stan. O código inclui os blocos:

- **data:** Neste bloco é definido o conjunto de dados.
- **parameters:** Aqui indicam-se os parâmetros, para os modelos propostos são o vetor de coeficientes de regressão β e o parâmetro de forma δ .
- **transformed parameters:** Nesta parte indicam-se os parâmetros transformados que serão calculados a partir do bloco de parâmetros como $\lambda = \log \delta$.
- **model:** No bloco de modelo são especificadas as prioris e a distribuição de probabilidade da variável resposta y .

```
data {  
  int<lower=0> k;  
  int<lower=0> n;  
  int<lower=0, upper=1> y[n];  
  matrix[n,k] X;  
}
```

```
parameters {
  real beta0;
  vector[k] beta;
  real loglambda;
}

transformed parameters {
  real lambda;
  vector[n] p;
  vector[n] prob;
  lambda = exp(loglambda);
  for (i in 1:n) {
    p[i] = Phi(beta0 + X[i]*beta);
    prob[i] = pow(p[i],lambda);
  }
}

model {
  beta0 ~ normal(0.0,100);
  beta ~ normal(0.0,100);
  loglambda ~ uniform(-2,2);
  y ~ bernoulli(prob);
}
```

B.2 Códigos no JAGS

A implementação dos códigos para JAGS é dada na linguagem BUGS, popularizada no uso do WinBUGS e OpenBUGS. A diferença da linguagem do Stan, aqui todos os elementos são inclusos no bloco do modelo.

```
model {
  for(i in 1:n) {
    y[i] ~ dbern(p[i])
    pl[i] <- phi(m[i])
    m[i] <- beta0 + X[i]*beta1
    p[i] <- pow(pl[i],lambda)
  }
}
```

```
beta0~ dnorm(0.0,0.01)
beta1 ~ dnorm(0.0,0.01)
loglambda ~ dunif(-2,2)
lambda<-exp(loglambda)
}
```

RESULTADOS DA SIMULAÇÃO

Neste anexo são apresentados os valores de viés, desvio padrão (DP), erro quadrático médio (EQM), tamanho efetivo de amostra (TEA), tempo (em segundos) e o valor da porcentagem de convergência (%C) segundo o critério de Geweke (GEWEKE, 1992). A porcentagem de convergência é a porcentagem de cadeias cujo valor z-score permaneceu entre os valores -2 e 2.

Nas Tabelas 22 a 31 são apresentadas as medidas em detalhe, considerando cada cenário de simulação, cada método de simulação e o tipo de cadeia 1 ou 2 (cadeia usual e cadeia menor).

Nas Figuras 34 a 39 são apresentados os vieses (em valor absoluto) dos parâmetros, segundo os tamanhos de amostra e os métodos de simulação NUTS e MHG, nos dois casos de simulação. Pode-se observar que enquanto o tamanho de amostra cresce, o viés diminui.

Tabela 22 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 50$ no caso 1 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
1	$\beta_0 = 0$	0,36	0,09	0,49	533,07		88	0,36	0,09	0,49	976,12		98
	$\beta_1 = 1$	0,16	0,06	0,23	469,27	93,07	92	0,16	0,06	0,23	975,73	8,18	90
	$\lambda = 1/3$	0,02	0,01	0,01	672,43		90	0,02	0,01	0,01	988,80		96
2	$\beta_0 = 0$	-0,91	0,07	1,11	478,68		94	-0,91	0,08	1,11	976,34		90
	$\beta_1 = 1$	0,90	0,13	1,67	435,63	82,05	96	0,88	0,13	1,56	983,12	8,36	94
	$\lambda = 1$	-0,64	0,01	0,42	651,93		92	-0,62	0,02	0,41	972,75		92
3	$\beta_0 = 0$	-5,10	1,73	172,70	347,78		96	-3,23	0,23	12,88	981,76		94
	$\beta_1 = 1$	3,60	1,74	160,59	314,17	80,72	94	1,80	0,24	6,01	976,36	8,43	96
	$\lambda = 3$	-2,61	0,02	6,82	614,88		96	-2,56	0,03	6,61	980,59		90

Tabela 23 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 200$ no caso 1 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
4	$\beta_0 = 0$	0,14	0,04	0,08	528,97		92	0,14	0,04	0,09	969,28		94
	$\beta_1 = 1$	0,05	0,02	0,03	509,09	142,74	92	0,05	0,02	0,03	972,90	32,64	90
	$\lambda = 1/3$	0,01	0,02	0,01	641,67		94	0,01	0,01	0,01	974,01		92
5	$\beta_0 = 0$	-0,98	0,06	1,12	522,91		88	-0,98	0,06	1,11	983,54		92
	$\beta_1 = 1$	0,49	0,04	0,31	491,66	174,95	90	0,49	0,04	0,31	977,53	33,24	94
	$\lambda = 1$	-0,58	0,04	0,41	637,72		90	-0,59	0,04	0,41	982,37		92
6	$\beta_0 = 0$	-2,42	0,09	6,27	482,81		94	-2,33	0,09	5,78	987,05		96
	$\beta_1 = 1$	0,90	0,05	0,93	461,02	146,11	94	0,86	0,05	0,85	982,89	33,80	90
	$\lambda = 3$	-2,43	0,06	6,11	612,35		96	-2,41	0,06	5,94	964,81		96

Tabela 24 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 500$ no caso 1 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
7	$\beta_0 = 0$	0,13	0,04	0,11	579,10		90	0,07	0,05	0,11	971,84		92
	$\beta_1 = 1$	-0,02	0,02	0,02	564,21	280,42	92	0,00	0,02	0,02	979,51	82,92	88
	$\lambda = 1/3$	0,07	0,04	0,07	657,18		90	0,04	0,03	0,06	989,22		92
8	$\beta_0 = 0$	-0,87	0,07	0,99	575,82		92	-0,85	0,07	1,00	984,74		88
	$\beta_1 = 1$	0,35	0,03	0,17	562,11	307,02	84	0,35	0,03	0,17	980,73	83,71	90
	$\lambda = 1$	-0,50	0,04	0,32	661,74		98	-0,45	0,05	0,33	988,77		90
9	$\beta_0 = 0$	-1,82	0,12	4,00	530,57		98	-1,79	0,12	3,91	968,80		98
	$\beta_1 = 1$	0,62	0,05	0,48	513,58	306,75	96	0,60	0,05	0,47	967,18	84,03	92
	$\lambda = 3$	-2,05	0,10	4,71	592,38		98	-2,02	0,11	4,63	966,00		98

Tabela 25 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 1000$ no caso 1 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
10	$\beta_0 = 0$	0,07	0,05	0,12	559,90		86	0,09	0,06	0,17	970,34		86
	$\beta_1 = 1$	0,01	0,02	0,02	542,41	544,96	90	0,00	0,02	0,02	976,97	166,34	90
	$\lambda = 1/3$	0,06	0,04	0,09	650,07		84	0,14	0,08	0,31	983,77		88
11	$\beta_0 = 0$	-0,77	0,10	1,05	624,36		92	-0,70	0,10	1,03	978,92		98
	$\beta_1 = 1$	0,31	0,04	0,16	611,73	578,75	96	0,28	0,04	0,16	974,50	169,49	98
	$\lambda = 1$	-0,32	0,08	0,44	690,78		92	-0,19	0,11	0,60	971,47		94
12	$\beta_0 = 0$	-1,36	0,14	2,74	572,84		94	-1,28	0,15	2,68	976,48		98
	$\beta_1 = 1$	0,44	0,05	0,30	571,01	606,85	98	0,42	0,05	0,30	973,08	169,35	96
	$\lambda = 3$	-1,54	0,16	3,61	622,75		94	-1,41	0,16	3,31	987,05		94

Tabela 26 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 2000$ no caso 1 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
13	$\beta_0 = 0$	0,01	0,06	0,16	627,05		86	0,00	0,06	0,17	971,81		96
	$\beta_1 = 1$	0,02	0,02	0,02	627,90	1147,97	98	0,02	0,02	0,02	973,32	372,00	90
	$\lambda = 1/3$	0,07	0,04	0,10	672,66		86	0,08	0,05	0,13	986,58		96
14	$\beta_0 = 0$	-0,35	0,10	0,60	711,44		88	-0,37	0,10	0,61	973,62		98
	$\beta_1 = 1$	0,13	0,03	0,08	707,75	1257,30	96	0,14	0,03	0,08	975,20	371,17	90
	$\lambda = 1$	0,03	0,11	0,57	706,54		86	0,03	0,11	0,60	971,12		94
15	$\beta_0 = 0$	-0,81	0,10	1,11	606,33		98	-0,78	0,12	1,32	972,69		92
	$\beta_1 = 1$	0,27	0,03	0,13	606,41	1379,18	90	0,26	0,04	0,15	969,02	387,44	98
	$\lambda = 3$	-1,09	0,14	2,12	640,97		98	-0,98	0,16	2,23	983,02		88

Tabela 27 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 50$ no caso 2 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
1	$\beta_0 = 0$	0,35	0,08	0,43	145,14		92	0,35	0,08	0,44	974,16		92
	$\beta_1 = 1$	0,22	0,06	0,22	141,58	81,87	96	0,21	0,05	0,17	975,43	1,75	84
	$\lambda = 1/3$	0,02	0,01	0,01	187,24		94	0,03	0,01	0,01	985,59		94
2	$\beta_0 = 0$	-1,10	0,08	1,53	128,25		88	-1,07	0,08	1,45	982,17		98
	$\beta_1 = 1$	0,94	0,13	1,65	125,37	76,61	92	0,91	0,12	1,51	977,07	1,74	78
	$\lambda = 1$	-0,64	0,01	0,42	178,76		98	-0,61	0,02	0,39	977,65		92
3	$\beta_0 = 0$	-3,49	0,22	14,58	93,32		90	-3,05	0,16	10,56	978,37		96
	$\beta_1 = 1$	1,91	0,22	5,98	85,66	76,98	98	1,62	0,16	3,89	977,77	1,75	84
	$\lambda = 3$	-2,64	0,01	6,98	175,41		90	-2,55	0,02	6,51	975,35		94

Tabela 28 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 200$ no caso 2 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
4	$\beta_0 = 0$	0,20	0,04	0,12	138,25		88	0,27	0,06	0,23	974,46		86
	$\beta_1 = 1$	0,04	0,02	0,02	135,52	88,38	96	0,01	0,02	0,02	975,24	6,65	94
	$\lambda = 1/3$	0,04	0,02	0,03	176,91		82	0,16	0,05	0,15	973,68		94
5	$\beta_0 = 0$	-0,96	0,07	1,18	136,72		88	-0,93	0,09	1,24	977,01		96
	$\beta_1 = 1$	0,42	0,04	0,25	133,62	87,97	94	0,41	0,05	0,27	961,81	6,77	94
	$\lambda = 1$	-0,53	0,05	0,38	179,22		88	-0,47	0,05	0,34	967,72		98
6	$\beta_0 = 0$	-2,41	0,12	6,48	121,12		86	-2,45	0,14	6,92	982,15		96
	$\beta_1 = 1$	0,97	0,08	1,28	119,31	90,61	98	0,97	0,08	1,27	980,06	6,90	84
	$\lambda = 3$	-2,44	0,06	6,13	165,92		88	-2,36	0,08	5,87	975,31		92

Tabela 29 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 500$ no caso 2 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
7	$\beta_0 = 0$	0,08	0,05	0,11	138,35		82	0,02	0,06	0,15	981,19		94
	$\beta_1 = 1$	0,01	0,02	0,01	137,34	130,15	92	0,03	0,02	0,02	976,10	16,84	86
	$\lambda = 1/3$	0,05	0,03	0,04	166,92		86	0,07	0,04	0,10	978,93		96
8	$\beta_0 = 0$	-0,73	0,08	0,87	156,21		96	-0,89	0,10	1,26	971,71		96
	$\beta_1 = 1$	0,30	0,03	0,14	154,68	128,23	98	0,35	0,04	0,20	974,29	17,09	98
	$\lambda = 1$	-0,34	0,08	0,41	190,78		94	-0,37	0,10	0,61	981,11		94
9	$\beta_0 = 0$	-2,07	0,11	4,89	144,07		88	-1,84	0,16	4,64	982,21		96
	$\beta_1 = 1$	0,72	0,05	0,62	143,03	123,81	92	0,64	0,06	0,58	977,44	17,11	88
	$\lambda = 3$	-2,26	0,08	5,38	172,64		84	-1,94	0,13	4,53	988,38		98

Tabela 30 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 1000$ no caso 2 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
10	$\beta_0 = 0$	0,13	0,06	0,19	153,31		86	0,20	0,08	0,39	970,88		92
	$\beta_1 = 1$	-0,02	0,02	0,02	147,19	178,86	96	-0,04	0,03	0,04	968,36	33,81	88
	$\lambda = 1/3$	0,14	0,05	0,16	190,04		86	0,42	0,13	1,01	981,55		92
11	$\beta_0 = 0$	-0,58	0,10	0,80	165,54		90	-0,67	0,13	1,22	970,55		98
	$\beta_1 = 1$	0,25	0,03	0,12	162,00	197,71	98	0,28	0,04	0,18	970,38	34,40	92
	$\lambda = 1$	-0,15	0,10	0,53	185,23		90	-0,05	0,15	1,07	966,86		96
12	$\beta_0 = 0$	-1,56	0,12	3,20	149,11		90	-1,58	0,17	3,89	975,98		88
	$\beta_1 = 1$	0,53	0,04	0,38	148,86	194,37	88	0,54	0,06	0,46	971,38	34,66	96
	$\lambda = 3$	-1,82	0,12	4,02	170,64		92	-1,68	0,16	4,08	980,06		92

Tabela 31 – Comparação de estimativas do modelo de regressão binária com ligação PP usando NUTS e MHG para tamanho de amostra $n = 2000$ no caso 2 de simulação.

Cenário	Parâmetro	NUTS						MHG					
		Viés	DP	EQM	TEA	Tempo	%C	Viés	DP	EQM	TEA	Tempo	%C
13	$\beta_0 = 0$	-0,02	0,05	0,12	165,72		92	0,02	0,07	0,27	977,14		100
	$\beta_1 = 1$	0,03	0,02	0,02	162,12	311,34	98	0,02	0,03	0,03	978,70	74,96	90
	$\lambda = 1/3$	0,02	0,03	0,04	197,66		90	0,16	0,09	0,40	988,57		98
14	$\beta_0 = 0$	-0,46	0,10	0,67	209,85		90	-0,50	0,12	0,96	974,62		98
	$\beta_1 = 1$	0,19	0,03	0,09	210,61	330,64	96	0,20	0,04	0,13	972,45	75,40	92
	$\lambda = 1$	-0,09	0,10	0,48	222,33		88	0,03	0,14	0,96	977,20		98
15	$\beta_0 = 0$	-1,09	0,12	1,93	172,34		94	-1,05	0,17	2,45	982,13		96
	$\beta_1 = 1$	0,36	0,04	0,21	175,03	346,76	96	0,35	0,05	0,26	982,12	76,08	96
	$\lambda = 3$	-1,28	0,18	3,26	172,97		96	-0,95	0,23	3,49	984,95		96

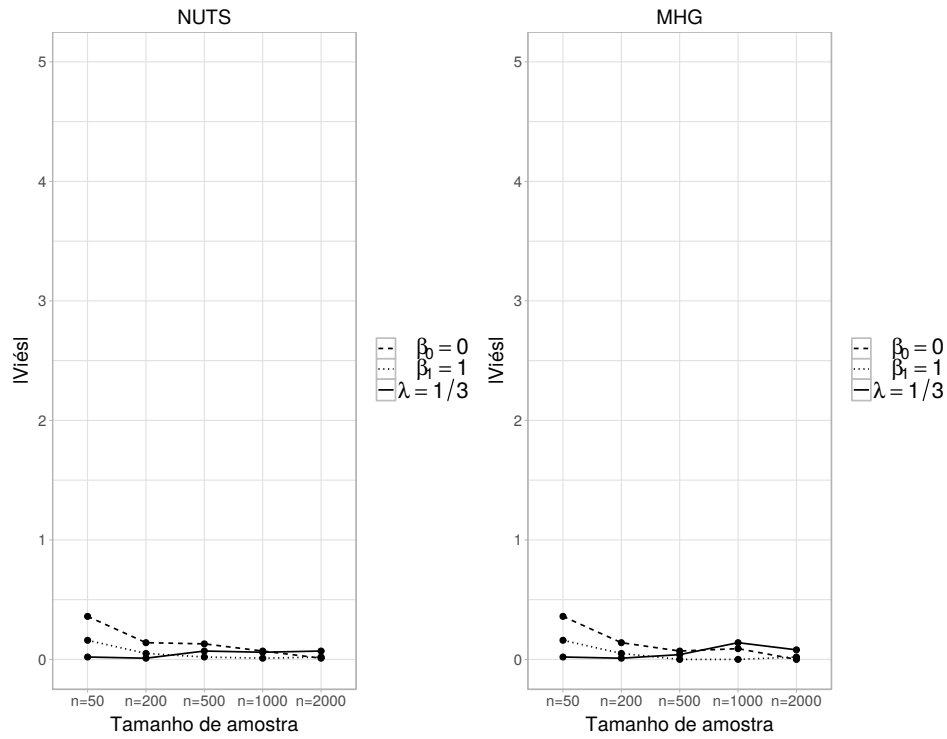


Figura 34 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 1/3$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 1 de simulação.

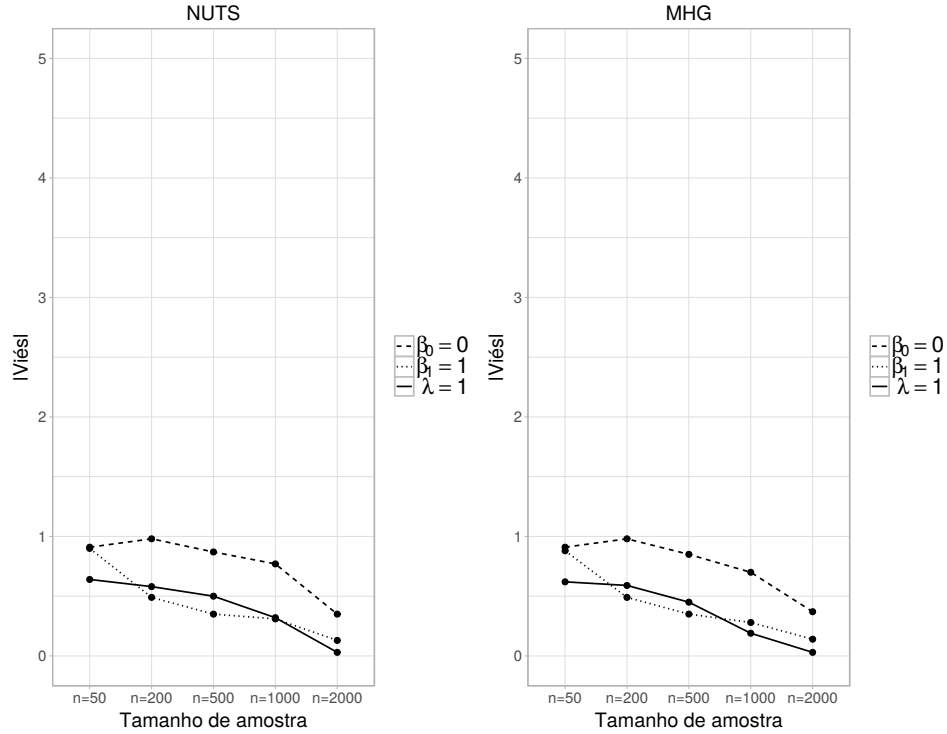


Figura 35 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 1$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 1 de simulação.

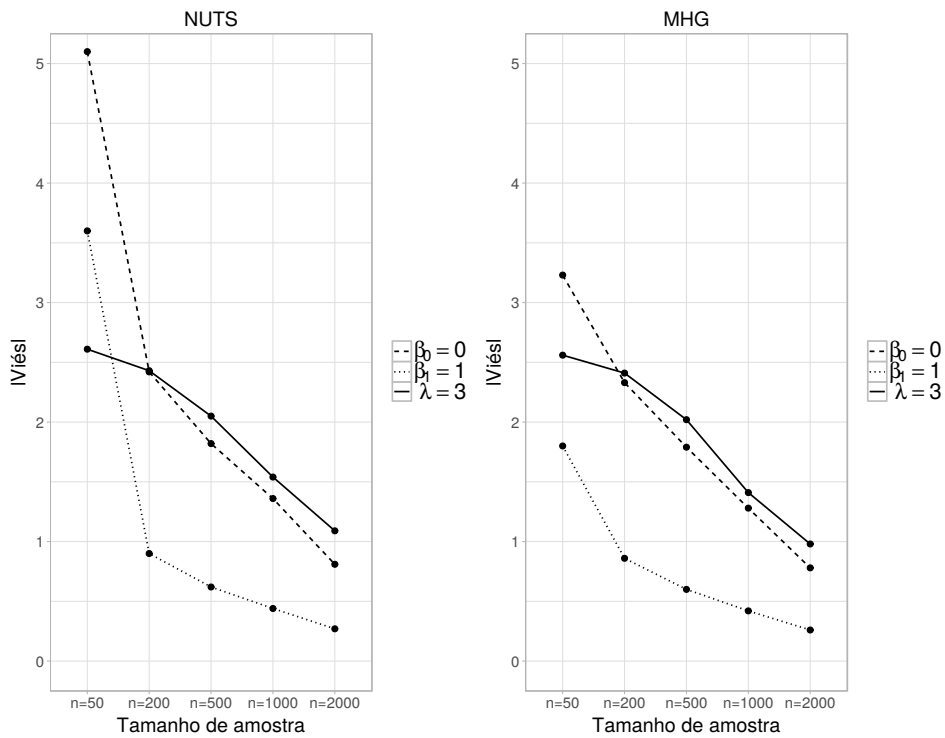


Figura 36 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 3$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 1 de simulação.

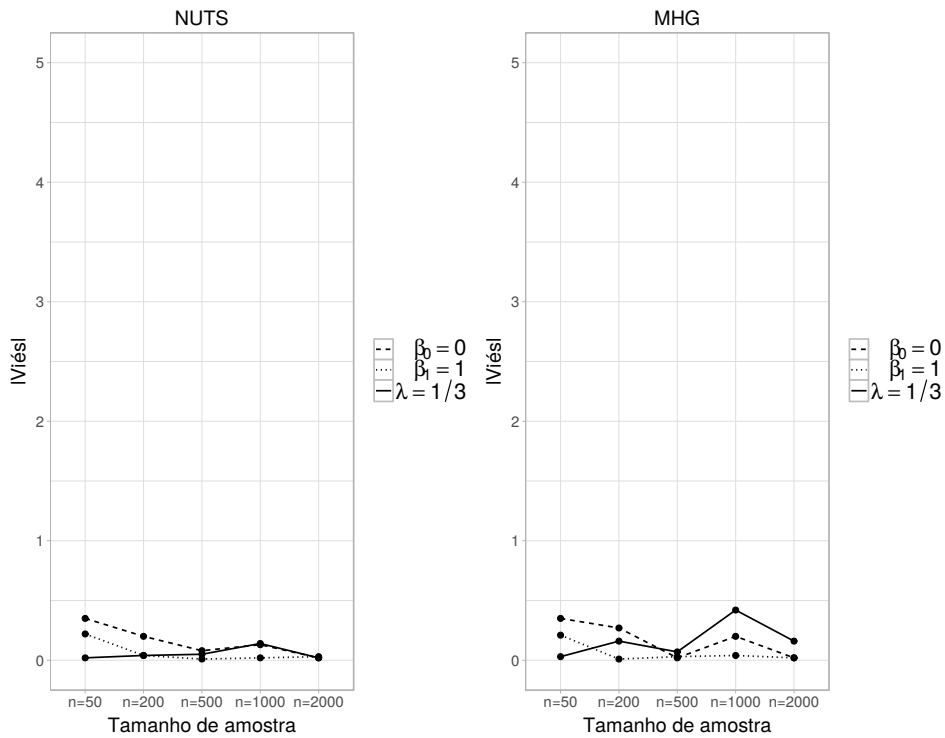


Figura 37 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 1/3$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 2 de simulação.

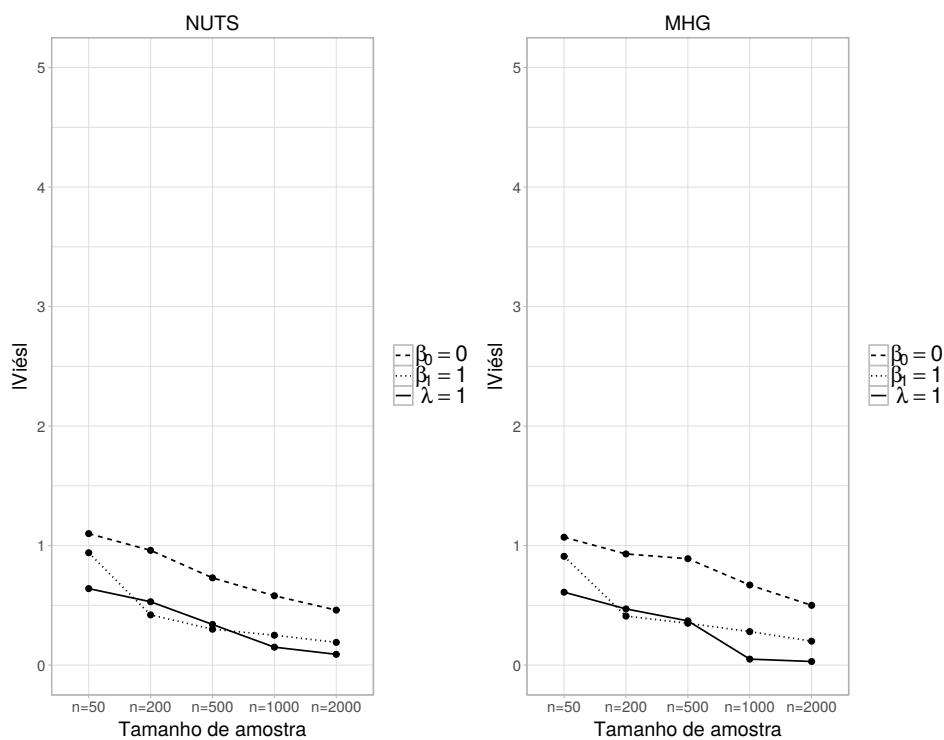


Figura 38 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 1$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 2 de simulação.

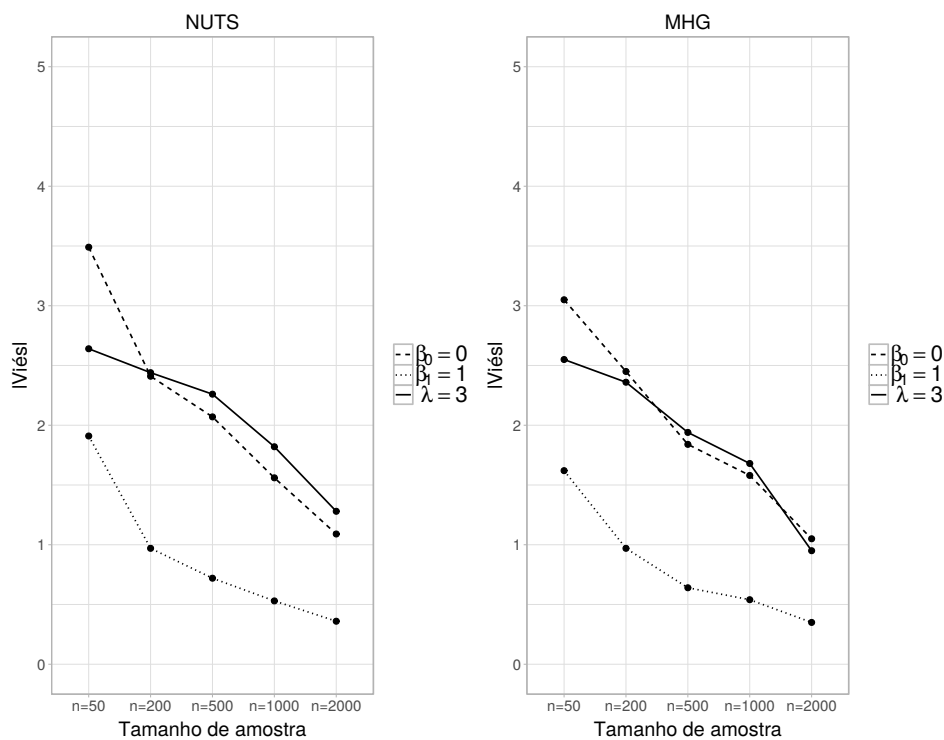


Figura 39 – Viés dos parâmetros $\beta_0 = 0, \beta_1 = 1, \lambda = 3$ segundo o tamanho de amostra, usando NUTS e MHG, no caso 2 de simulação.

RESULTADOS DA APLICAÇÃO

Neste anexo são apresentadas as medidas de avaliação preditiva e os resultados do modelo de regressão binária com função de ligação potência logit, na aplicação a dados educacionais.

D.1 Classificações segundo pontos de corte para a aplicação

As Tabelas 32 a 35 incluem as classificações segundo os diversos pontos de corte e as medidas Verdadeiro positivo (VP), Falso negativo (FN), Falso positivo (FP) e Verdadeiro negativo (VN) vistas no Capítulo 3, para o modelo de regressão binária com as quatro funções de ligação inicialmente escolhidas como as melhores.

Tabela 32 – Frequência das classificações segundo diferentes pontos de corte para o modelo de regressão binária com ligação **PL**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Classificações	VP	1075	853	634	424	285	133	39	14	0
	FP	2793	1437	755	416	205	71	25	4	0
	FN	248	470	689	899	1038	1190	1284	1309	1323
	VN	9143	10499	11181	11520	11731	11865	11911	11932	11936

Tabela 33 – Frequência das classificações segundo diferentes pontos de corte para o modelo de regressão binária com ligação **RPL**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Classificações	VP	1075	863	667	424	240	78	17	0	0
	FP	2782	1513	810	416	171	41	8	0	0
	FN	248	460	656	899	1083	1245	1306	1323	1323
	VN	9154	10423	11126	11520	11765	11895	11928	11936	11936

Tabela 34 – Frequência das classificações segundo diferentes pontos de corte para o modelo de regressão binária com ligação **PP**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Classificações	VP	1075	838	611	424	296	159	77	17	0
	FP	2776	1366	703	416	216	84	39	9	0
	FN	248	485	712	899	1027	1164	1246	1306	1323
	VN	9160	10570	11233	11520	11720	11852	11897	11927	11936

Tabela 35 – Frequência das classificações segundo diferentes pontos de corte para o modelo de regressão binária com ligação **RPP**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Classificações	VP	1078	848	611	415	282	133	49	14	0
	FP	2815	1415	703	405	200	70	27	4	0
	FN	245	475	712	908	1041	1190	1274	1309	1323
	VN	9121	10521	11233	11531	11736	11866	11909	11932	11936

D.2 Medidas de avaliação preditiva para a aplicação

A partir das classificações segundo os pontos de corte é possível obter as medidas de avaliação preditiva, dadas também considerando diversos pontos de corte, nas Tabelas 36 a 39, para todos esses modelos. Nas Figuras 40 a 43 também podem ser vistas as medidas de avaliação preditiva para cada um dos modelos.

Tabela 36 – Medidas de avaliação preditiva segundo diferentes pontos de corte para o modelo de regressão binária com ligação **PL**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Medidas	RVP	0,813	0,645	0,479	0,320	0,215	0,101	0,029	0,011	0
	RFN	0,187	0,355	0,521	0,680	0,785	0,899	0,971	0,989	1
	RFP	0,234	0,120	0,063	0,035	0,017	0,006	0,002	0	0
	RVN	0,766	0,880	0,937	0,965	0,983	0,994	0,998	1	1

Tabela 37 – Medidas de avaliação preditiva segundo diferentes pontos de corte para o modelo de regressão binária com ligação **RPL**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Medidas	RVP	0,813	0,652	0,504	0,320	0,181	0,059	0,013	0	0
	RFN	0,187	0,348	0,496	0,680	0,819	0,941	0,987	1	1
	RFP	0,233	0,127	0,068	0,035	0,014	0,003	0,001	0	0
	RVN	0,767	0,873	0,932	0,965	0,986	0,997	0,999	1	1

Tabela 38 – Medidas de avaliação preditiva segundo diferentes pontos de corte para o modelo de regressão binária com ligação **PP**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Medidas	RVP	0,813	0,633	0,462	0,320	0,224	0,120	0,058	0,013	0
	RFN	0,187	0,367	0,538	0,680	0,776	0,880	0,942	0,987	1
	RFP	0,233	0,114	0,059	0,035	0,018	0,007	0,003	0,001	0
	RVN	0,767	0,886	0,941	0,965	0,982	0,993	0,997	0,999	1

Tabela 39 – Medidas de avaliação preditiva segundo diferentes pontos de corte para o modelo de regressão binária com ligação **RPP**.

		Pontos de corte								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Medidas	RVP	0,815	0,641	0,462	0,314	0,213	0,101	0,037	0,011	0
	RFN	0,185	0,359	0,538	0,686	0,787	0,899	0,963	0,989	1
	RFP	0,236	0,119	0,059	0,034	0,017	0,006	0,002	0	0
	RVN	0,764	0,881	0,941	0,966	0,983	0,994	0,998	1	1

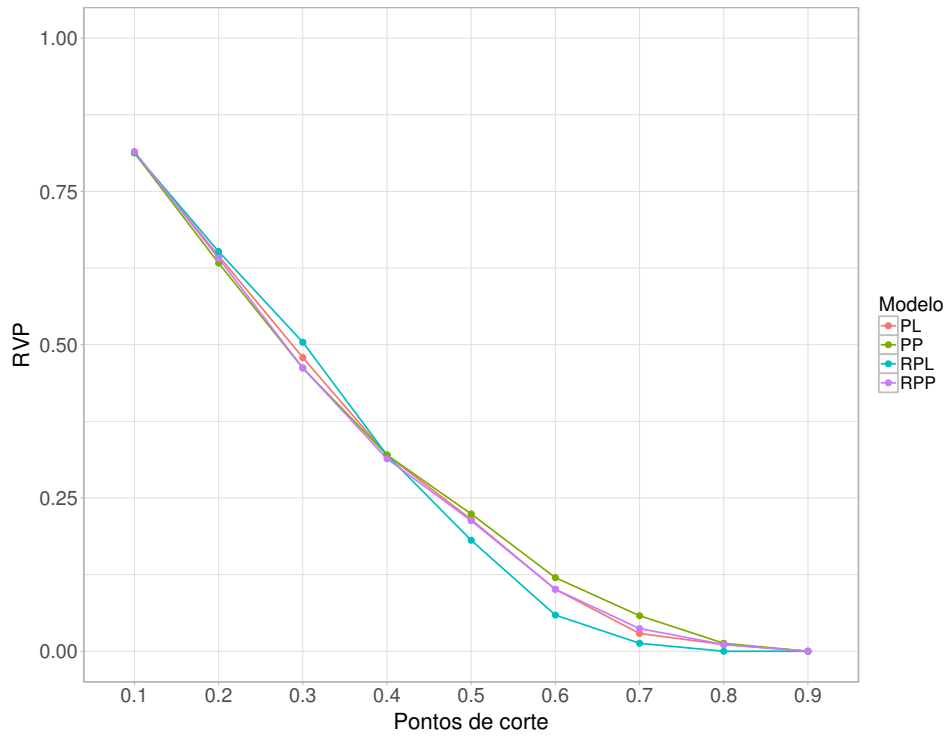


Figura 40 – Sensibilidade ou razão de verdadeiro positivo.

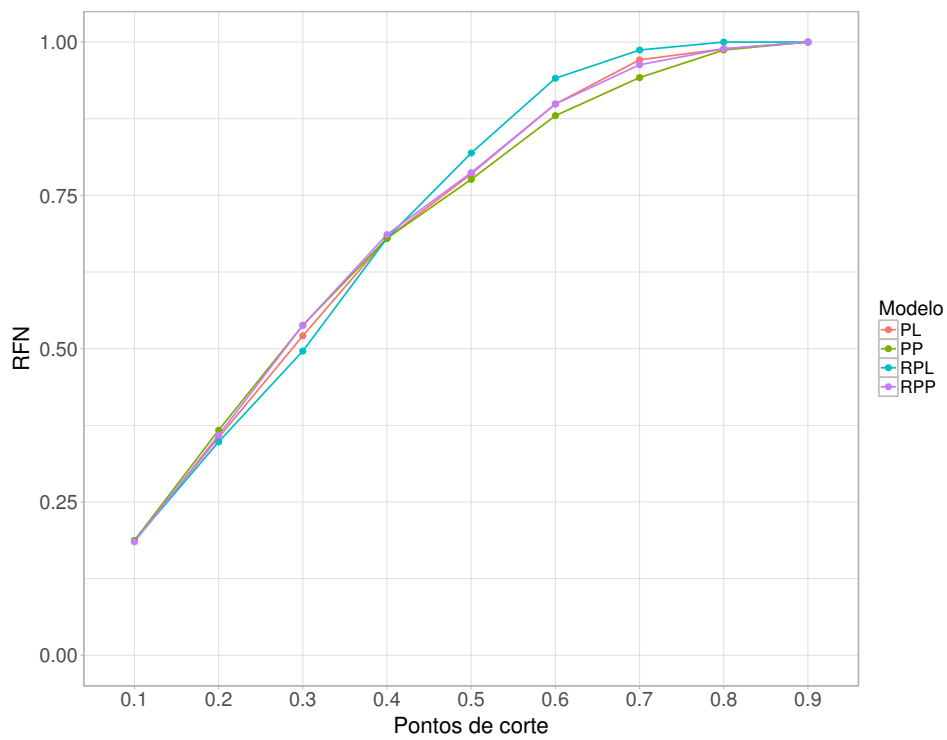


Figura 41 – Razão de falso negativo.

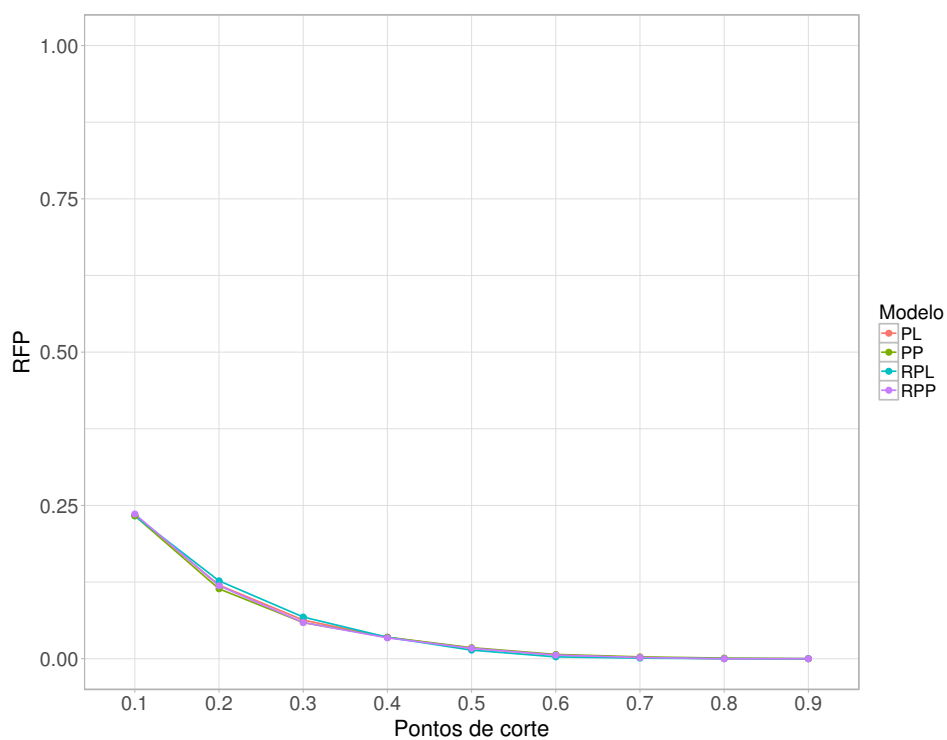


Figura 42 – Razão de falso positivo.

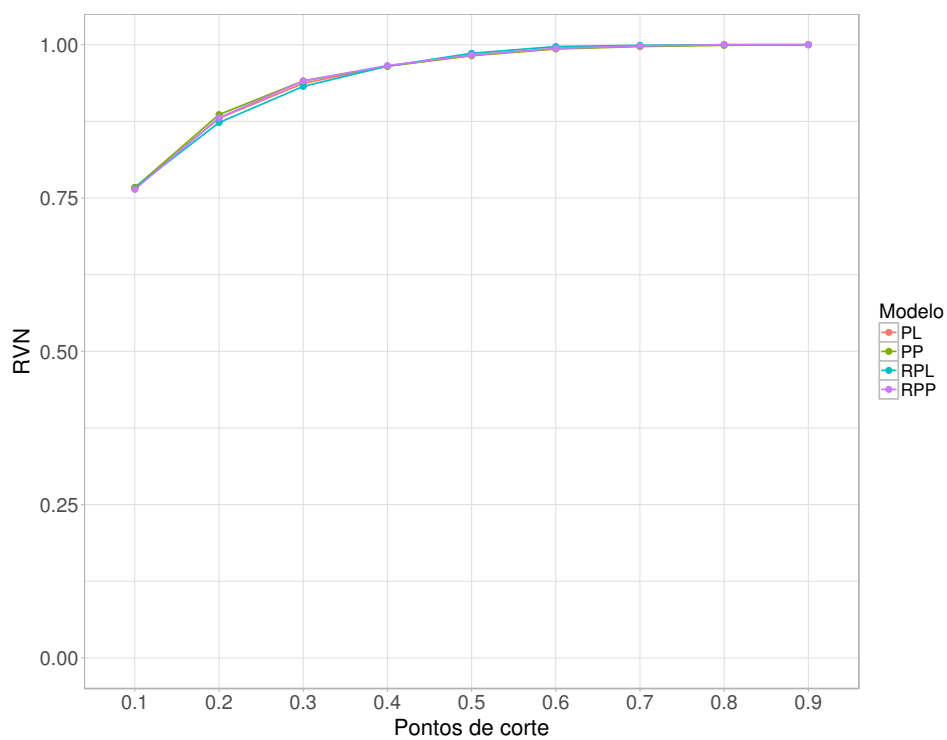


Figura 43 – Especificidade ou razão de verdadeiro negativo.

D.3 Modelo de regressão binária com função de ligação potência logit

Na Tabela 40 é apresentado o resumo dos parâmetros considerando a ligação potência logit (PL). Observa-se que nenhum dos intervalos de credibilidade de 95% inclui o valor zero, e todos os parâmetros são significativos. Comparando os resultados com o modelo de regressão binária com função de ligação reversa de potência logit (RPL), é visto que o sinal para cada um dos coeficientes das covariáveis é mantido, dessa forma, a interpretação é mantida. A diferença encontra-se no valor do parâmetro $\lambda = 4,88$, em comparação ao valor do parâmetro do modelo com função de ligação RPL, $\lambda = 0,19$. O valor do parâmetro λ era esperado em cada um dos modelos, segundo a proporção de 1's na variável resposta, perto de 10%.

Tabela 40 – Resumo a posteriori dos parâmetros do modelo com função de ligação PL.

Variáveis	Parâmetros	média	DP	mediana	IC _{95%}	
Intercepto	β_0	-3,23	0,53	-3,22	-4,36	-2,38
Zona	β_1	-0,28	0,07	-0,27	-0,43	-0,15
Gestão	β_2	0,57	0,05	0,57	0,48	0,67
Sexo	β_3	0,24	0,04	0,24	0,17	0,33
Língua espanhola	β_4	0,01	0,00	0,01	0,01	0,01
Assimetria	δ	1,56	0,25	1,55	1,05	1,97
	λ	4,88	1,20	4,73	2,87	7,20

DP: Desvio padrão. IC: Intervalo de credibilidade.