

**Modelos Bayesianos semiparamétricos  
para dados binários**

Márcio Augusto Diniz

TESE APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Carlos Alberto de Bragança Pereira

Coorientador: Adriano Polpo de Campos

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES/CNPq

São Paulo, 11 Agosto de 2015

# Modelos Bayesianos semi-paramétricos para dados binários

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 11/06/2015. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Carlos Alberto de Bragança Pereira (orientador), IME-USP
- Prof. Dr. Adriano Polpo de Campos (coorientador), UFSCar
- Prof. Dr. Ronaldo Dias, UNICAMP
- Prof. Dr. Claudio José Struchiner, FIOCRUZ
- Prof. Dr. Eduardo Massad, FM-USP

# Agradecimentos

Há dez anos atrás, eu comecei o estudo da ciência estatística. Naquele tempo, jamais imaginei quão alto eu poderia voar.

Desta forma, eu gostaria de agradecer a todos envolvidos na construção do colégio Engenheiro Juarez Wanderley, pois sem tal projeto eu não poderia ter tido a oportunidade de cursar uma universidade pública. Agradeço aos professores e meus amigos daquele tempo.

Alguns deles ainda estão presentes em minha vida diária: agradeço à Érica Marques por ser minha irmã em São Paulo mesmo sem laços sanguíneos e toda sua família; à Kenya Marcon pelas correções deste texto e por elucidar tantas verdades sobre o mundo; ao Jaime Teles por ter me estimulado a uma vida mais ativa no início do doutorado e ao Thiago Juarez por sempre me apresentar bons jogos que tornam os meus dias mais leves.

Em memória, meu grande companheiro Thiago Carneiro que se tornou uma estrela a iluminar nossas vidas mais cedo do que eu gostaria e espero que esteja me assistindo lá do céu.

Nesta longa jornada, meus primeiros anos no estudo da ciência estatística foram na Universidade Estadual de Campinas. Foi lá que eu me apaixonei pela estatística e descobri que queria ser cientista.

Agradeço à professora Laura Rifo que foi a responsável pelo meu deslumbramento pela estatística e os meus amigos que foram meus companheiros nos primeiros e difíceis anos desta jornada. Em especial, Denny Mineki que sempre acreditou em mim mais do que eu durante toda a graduação e me ensinou a ter serenidade e perseverança frente aos desafios; à Beatriz Cuyabano e Guilherme Ludwig com quem compartilhei os desafios do mestrado; ao casal gaúcho mais querido que conheço, Karen Jung e Marcio Valk e a gaúcha mais determinada e divertida que tive a honra de conviver, Lorena Vicini.

Agradeço ao meu primeiro mentor e amigo Luiz Koodi Hotta que me adotou no terceiro ano de graduação e me ajudou a descobrir a minha vocação em ser um pesquisador.

Como último grande desafio (eu acreditava), fui para a cidade grande a fim de estudar e me tornar um doutor em ciências. Eu não poderia ter sido mais sortudo ao ter como mentores Carlos Alberto de Bragança Pereira e Adriano Polpo de Campos. Obrigado pela orientação e iluminar meu caminho para me tornar um pesquisador independente, pelo indiscreto carinho e por me proporcionar desafios ainda maiores após o doutorado. Espero que eu possa ser um ser humano tão grande como vocês no futuro.

Agradeço ao Júlio Pereira e José Jukemura por me desafiarem a ser um estatístico além do convencional e buscar constantemente a ligação entre teoria e aplicação da estatística. Obrigado aos colegas de trabalho que tornavam os dias mais fáceis e menos estressantes no Hospital das Clínicas. Em especial, Antônio Turri, Mariza Ochner, Priscila Caproni, Fabiana Soares, Myrthes Freire e Viviane Franceschet.

Ao lado do desafio do meu doutorado, eu aceitei um outro desafio nestes últimos anos: aprender

natação. Obrigado ao Alexandre Vieira, Flavia Sousa e Raphael Pabst por pacientemente ensinarem uma pessoa tão descoordenada como eu em uma atividade tão diferente da minha rotina diária que me permitiu renovar minhas energias frente aos difíceis momentos da pesquisa.

Eu também pude construir grandes laços nestes últimos anos e conviver com bons amigos na vida acadêmica que me ensinaram inúmeras lições: Aldo Medina, Francielle Lima, Catarina Azeredo, Juliana Yukari, Tiago Magalhães, Michel Helcias, João Ítalo, Alice Lemos, Fabio Oki, Mel Oliveira, Leandro Tavares, Vinicius Calsavara, Agatha Rodrigues, Brian Melo e Raony Cassab.

Agradeço à minha família pela compreensão e extraordinário suporte ao longos de todos esses anos: meus pais Maria e Antônio; meus irmãos Marco, Maurício e Márcia e meus sobrinhos Daniel e Vivian. São as pessoas mais especiais na minha vida.

Finalmente, agradeço à banca pela paciência na leitura deste texto e pelas gentis correções assim como as agências CAPES e CNPq pelo suporte financeiro.

# Resumo

DINIZ, M. A. **Modelos Bayesianos semi-paramétricos para dados binários**. 2015. 66 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015. Este trabalho propõe modelos Bayesianos semi-paramétricos para dados binários. O primeiro modelo é uma mistura em escala que permite lidar com discrepâncias relacionadas a curtose do modelo *Logístico*. É uma extensão relevante a partir do que já foi proposto por [Basu e Mukhopadhyay \[1998\]](#) ao possibilitar a interpretação da distribuição a priori dos parâmetros através de razões de chances.

O segundo modelo usufrui da mistura em escala em conjunto com a transformação proposta por [Yeo e Johnson \[2000\]](#) possibilitando que a curtose assim como a assimetria sejam ajustadas e um parâmetro informativo de assimetria seja estimado. Esta transformação é muito mais apropriada para lidar com valores negativos do que a transformação de [Box e Cox \[1964\]](#) utilizada por [Guerrero e Johnson \[1982\]](#) e é mais simples do que o modelo proposto por [Stukel \[1988\]](#).

Por fim, o terceiro modelo é o mais geral entre todos e consiste em uma mistura de posição e escala tal que possa descrever curtose, assimetria e também bimodalidade. O modelo proposto por [Newton \*et al.\* \[1996\]](#), embora, seja bastante geral, não permite uma interpretação palpável da distribuição a priori para os pesquisadores da área aplicada.

A avaliação dos modelos é realizada através de medidas de distância de probabilidade Cramér-von Mises, Kolmogorov-Smirnov e Anderson-Darling e também pelas Ordenadas Preditivas Condicionais.

**Palavras-chave:** Bayesiana semi-paramétrico, dados binários, assimetria, regressão logística



# Abstract

DINIZ, M. A. **Bayesian semi-parametric models for binary data**. 2015. 66 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

This work proposes semi-parametric Bayesian models for binary data. The first model is a scale mixture that allows handling discrepancies related to kurtosis of *Logistic* model. It is a more interesting extension than has been proposed by Basu e Mukhopadhyay [1998] because this model allows the interpretation of the prior distribution of parameters using odds ratios.

The second model enjoys the scale mixture together with the scale transformation proposed by Yeo e Johnson [2000] modeling the kurtosis and the asymmetry such that a parameter of asymmetry is estimated. This transformation is more appropriate to deal with negative values than the transformation of Box e Cox [1964] used by Guerrero e Johnson [1982] and simpler than the model proposed by Stukel [1988].

Finally, the third model is the most general among all and consists of a location-scale mixture that can describe kurtosis and skewness also bimodality. The model proposed by Newton *et al.* [1996], although general, does not allow a tangible interpretation of the a priori distribution for researchers of applied area.

The evaluation of the models is performed through distance measurements of distribution of probabilities Cramer-von Mises Kolmogorov-Smirnov and Anderson-Darling and also the Conditional Predictive sorted.

**Keywords:** Bayesian Semi-parametric, binary data, asymmetry, logistic regression



# Sumário

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Modelos de Regressão Binária</b>	<b>5</b>
2.1 Processos de Dirichlet . . . . .	6
2.2 O primeiro modelo: mistura em escala . . . . .	8
2.2.1 <i>Gama</i> . . . . .	9
2.2.2 <i>Kolmogorov-Smirnov</i> . . . . .	9
2.2.3 <i>Exponencial</i> . . . . .	10
2.3 Transformações Potência . . . . .	10
2.4 O segundo modelo: transformação potência . . . . .	12
2.5 O terceiro modelo: mistura de posição e escala . . . . .	12
2.6 Elicitação da distribuição a priori . . . . .	13
2.7 Distribuição a posteriori . . . . .	15
2.8 Estimação pontual de $\beta$ , $\lambda$ e $F$ . . . . .	16
2.9 Distribuição Preditiva . . . . .	16
2.10 Ordenada Preditiva Condicional . . . . .	17
2.11 Distâncias entre medidas de probabilidade . . . . .	17
<b>3 Exemplos Simulados</b>	<b>19</b>
3.1 Modelo Logístico . . . . .	20
3.2 Modelo Log-log . . . . .	22
3.3 Modelo de Mistura <i>Logístico</i> . . . . .	26
<b>4 Aplicações</b>	<b>29</b>
4.1 Mortalidade dos Besouros . . . . .	29
4.2 Idade da primeira menstruação . . . . .	33
<b>5 Algoritmo: mistura em escala</b>	<b>37</b>
5.1 Mistura em escala: variáveis latentes . . . . .	37
5.2 Amostradores de Gibbs . . . . .	38
5.2.1 Amostragem de $\mathbf{Z}_\tau   \mathbf{W}, \mathbf{X}, \beta, \mathbf{K}, v$ . . . . .	39
5.2.2 Amostragem de $\mathbf{K}   \mathbf{W}, \mathbf{X}, \beta, \mathbf{Z}, \mathbf{q}$ . . . . .	42

5.2.3	Amostragem de $\mathbf{q} \mathbf{K}, \alpha$ . . . . .	43
5.2.4	Amostragem de $\beta \mathbf{W}, \mathbf{X}, \mathbf{Z}, \mathbf{K}, \mu_\beta, \mathbf{T}_\beta$ . . . . .	43
5.2.5	Amostragem de $\mathbf{W} \mathbf{Y}, \mathbf{X}, \beta, \mathbf{Z}_\tau, \mathbf{K}$ . . . . .	44
5.2.6	Algoritmo . . . . .	44
5.3	Ilustração comparativa dos algoritmos . . . . .	45
<b>6</b>	<b>Considerações Finais</b>	<b>49</b>
	<b>Referências Bibliográficas</b>	<b>51</b>

# Lista de Figuras

3.1	Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo <i>Logístico</i> com 30 ensaios <i>Bernouli</i> . . . . .	22
3.2	Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo <i>Log-log</i> com 30 ensaios <i>Bernouli</i> . . . . .	25
3.3	Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo de Mistura <i>Logístico</i> com 30 ensaios <i>Bernouli</i> . . . . .	28
4.1	Estimativas das curvas de probabilidade do conjunto de dados Mortalidade dos Be-souros . . . . .	32
4.2	Estimativas das curvas de probabilidade do conjunto de dados Idade da primeira menstruação . . . . .	35
5.1	Histórico das cadeias de Markov para os pesos do processo <i>Dirichlet</i> do modelo MEVL	47
5.2	Histórico das cadeias de Markov para os pesos do processo <i>Dirichlet</i> do modelo ME	48



# Lista de Tabelas

2.1	Estudo Retrospectivo . . . . .	14
3.1	Dados do modelo <i>Logístico</i> com 10 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	20
3.2	Dados do modelo <i>Logístico</i> com 10 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	21
3.3	Dados do modelo <i>Logístico</i> com 30 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	21
3.4	Dados do modelo <i>Logístico</i> com 30 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	21
3.5	Dados do modelo <i>Log-log</i> com 10 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	23
3.6	Dados do modelo <i>Log-log</i> com 10 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	23
3.7	Dados do modelo <i>Log-log</i> com 30 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	24
3.8	Dados do modelo <i>Log-log</i> com 30 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	24
3.9	Dados do modelo de Mistura <i>Logístico</i> com 10 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	26
3.10	Dados do modelo <i>Logístico</i> com 10 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	27
3.11	Dados do modelo de Mistura <i>Logístico</i> com 30 ensaios <i>Bernouli</i> - Estimativas para os modelos . . . . .	27
3.12	Dados do modelo de Mistura <i>Logístico</i> com 30 ensaios <i>Bernouli</i> - Medidas de distância e redução percentual relativa . . . . .	27
4.1	Conjunto de dados: Mortalidade dos besouros . . . . .	29
4.2	Mortalidade dos besouros - Condições para o algoritmo MCMC . . . . .	30
4.3	Mortalidade dos besouros - Estimativas para os modelos . . . . .	31
4.4	Mortalidade dos besouros - Medidas de distância e redução percentual relativa . . . . .	31
4.5	Conjunto de dados: Idade da primeira menstruação . . . . .	33
4.6	Idade da primeira menstruação - Condições para o algoritmo MCMC em modelos <i>Logístico</i> . . . . .	33
4.7	Idade da primeira menstruação - Estimativas para os modelos bayesianos <i>Logístico</i> . . . . .	34
4.8	Idade da primeira menstruação - Medidas de distância e redução percentual relativa para os modelos <i>Logístico</i> . . . . .	34

5.1	Ilustração comparativa dos algoritmos - Condições para o algoritmo MCMC em modelos <i>Logístico</i> . . . . .	45
5.2	Ilustração comparativa dos algoritmos - Estimativas para os modelos bayesianos <i>Logístico</i> . . . . .	45
5.3	Ilustração comparativa dos algoritmos - Medidas de distância e redução percentual relativa para os modelos <i>Logístico</i> . . . . .	45

# Capítulo 1

## Introdução

A literatura sobre dados binários é bastante vasta e há uma enorme variedade de modelos sob a perspectiva clássica e bayesiana. Ainda assim, o modelo de regressão *Logística* continua sendo extensivamente utilizado, independente da qualidade do ajuste, em diversas áreas de pesquisa aplicada.

Antes de discutir os trabalhos recentes sobre o assunto e estabelecer o contexto sobre o qual este texto pretende contribuir, é importante entender como o modelo *Logístico* tornou-se uma ferramenta central para a modelagem de dados binários e, desta forma, uma breve revisão histórica baseada em Cramer [2011] será feita.

O mais antigo registro corresponde a Alphonse Quetelet (1795 - 1874), um astrônomo belga que tornou-se estatístico e questionou a suposição de crescimento exponencial defendida por Malthus [1872], pois, poderia levar a valores não factíveis de um tamanho de população. Quetelet apresentou esta preocupação para seu aluno Pierre-François Verhulst (1804 - 1849), matemático, que adicionou um termo para representar a existência de um limitante ao crescimento indiscriminado na taxa de crescimento.

O problema formulado no contexto de equações diferenciais tinha como solução justamente a função que foi denominada *Logística*. Tais ideias foram publicadas em três artigos no período entre 1838 e 1847 com exemplos de crescimento populacional de países como França, Bélgica e Rússia até 1833.

Na década de 20, Raymond Pearl (1879 - 1940) e Lowell J. Reed (1886 - 1966) redescobriram o modelo *Logístico* estudando o crescimento populacional frente a limitação de disponibilidade de alimentos criada devido a Primeira Guerra Mundial, entretanto, sem nomeá-la como Verhulst. Somente em 1925, George Yule (1871 - 1951) recuperou o termo proposto por Verhulst.

Ao mesmo tempo, o modelo de regressão *Probit* surgia no artigo de Bliss [1934] no contexto de experimentos dose-resposta em que a dose é fixa e a resposta é aleatória frente a variabilidade individual em torno da média de tolerância. O termo *Probit* foi introduzido também por Bliss [1935] como uma sigla para a medida de unidade de probabilidade que seria uma forma mais conveniente de expressar desvios da média de uma distribuição *Normal*.

A popularidade do modelo de regressão *Probit* no início do século 20 era significativa e Joseph Berkson (1899 - 1922), estatístico chefe da clínica Mayo, empreendeu um grande esforço a fim de popularizar a regressão *Logística*. Esta se sobressaiu nos dois métodos de estimação mais populares (Mínimos Quadrados e Máxima Verossimilhança) frente a inexistência do computador, porém, o modelo *Probit* possibilitava a interpretação em desvios da média de uma distribuição *Normal*, o que parecia ser essencial para os estudos de dose-resposta.

Ao longo das décadas de 60 e 70, outras aplicações da regressão *Logística* foram descobertas: modelos log-lineares e estudos retrospectivos. Cox [1969] é um trabalho bastante representativo em direção a popularização deste modelo.

Nelder e Weddeburn [1972] introduziram os Modelos Lineares Generalizados e demonstraram que a função de ligação definida pela função de distribuição *Logística* é a escolha natural para estabelecer uma relação linear entre a probabilidade de sucesso e as covariáveis.

Por estes e tantos outros motivos, o modelo *Logístico* tornou-se padrão nos pacotes computacionais disponíveis e é frequentemente o único modelo utilizado. Por outro lado, tanto o modelo *Probit* como o modelo *Logístico* são simétricos, isto é, a probabilidade de sucesso aproxima-se do valor um na mesma taxa que se aproxima do valor zero. Esta suposição pode não ser válida e ignorá-la pode gerar conclusões errôneas.

Frente a esta preocupação, as alternativas iniciais são o modelo *Log-log Complementar* com assimetria a direita e o modelo *Log-log* com assimetria a esquerda provenientes das distribuições *Valor Extremo* para o mínimo e máximo, respectivamente.

Na abordagem clássica, Aranda-Ordaz [1981] introduziu uma transformação sobre a probabilidade de sucesso no modelo *Logístico* tal que este fosse um caso específico assim como o modelo *Log-log Complementar*. Prentice [1976] sugere uma função de ligação baseada na função de distribuição *F-Snedecor* em que as funções de ligação *Logito*, *Log-log Complementar*, *Log-log*, *Probit*, *Laplace* e *Exponencial* são casos particulares.

Na mesma corrente de ideias, Guerrero e Johnson [1982] recorrem a transformação potência de Box e Cox [1964] e Stukel [1988] baseia-se na transformação definida por Manly [1976] para generalizar a regressão *Logística* e lidar com discrepâncias de assimetria e curtose. Caron e Polpo [2009] apresentam o modelo de regressão baseado na função de distribuição *Weibull* que pode aproximar as funções de ligação *Log-log Complementar*, *Probit* e *Logito*.

Na perspectiva bayesiana, Albert e Chib [1993] constroem o modelo *t-student* baseado na mistura em escala de componentes com distribuição *Normal* através da introdução de variáveis latentes como um artifício para a redução do problema ao contexto de regressão linear e sugerem uma aproximação para o modelo *Logístico*.

A partir de tais ideias, Czado [1994] utiliza a transformação potência de Box e Cox [1964] para modelar assimetria aplicando sobre o preditor linear, Chen e Dey [1998] consideram dados binários correlacionados e também indicam como escrever a distribuição *Logística* através de uma mistura em escala de componentes *Normal* com a distribuição de *Kolmogorov-Smirnov*.

Chen *et al.* [1999] sugerem uma mistura de locação para descrever curvas de probabilidade assimétricas que mais tarde foi definida por Sahu *et al.* [2003] como uma versão alternativa da distribuição *Normal Assimétrica*. Bazan *et al.* [2010] generalizam a abordagem para considerar também a distribuição *Normal Assimétrica* definida por Azzalini [1985].

Newton *et al.* [1996] propõem uma classe de funções de distribuição sem restrições além da monotocidade e parâmetros de posição e escala conhecidos que em conjunto com uma distribuição a priori dada por um processo *Dirichlet* modificado possibilitam um modelo bayesiano semi-paramétrico.

Nesta mesma linha, Basu e Mukhopadhyay [1998] através de misturas em escala da distribuição *Normal* e o algoritmo Amostrador de Gibbs por Urna de Pólia [Escobar, 1994] sugerem a versão bayesiana semi-paramétrica das ideias propostas por Albert e Chib [1993], contudo, limitada ao modelo *Probit* a priori devido a dificuldade do Amostrador de Gibbs por Urna de Pólia em lidar com distribuições não conjugadas.

Por fim, Polson *et al.* [2013] discutem um novo conjunto de variáveis latentes baseadas na distribuição *Polia-Gama* que permite aproximar a regressão *Logística* através de uma mistura em locação de distribuições *Normal* no contexto bayesiano paramétrico.

É visível que há um grande esforço da estatística bayesiana em formular soluções para a modelagem de dados binários através da regressão *Logística* devido a grande popularidade desta ferramenta e também a interpretação dos coeficientes de regressão em termos de razões de chances.

A razão de chance é uma medida de efeito típica de estudos retrospectivos e também é a interpretação natural dos coeficientes de regressão no contexto de Modelos Lineares Generalizados, por isso, há tanta dificuldade em utilizar outros modelos dentro da pesquisa aplicada mesmo quando o modelo *Logístico* não é apropriado, isto é, há discrepâncias de curtose, assimetria e bimodalidade da distribuição *Logística*.

Frente a tais considerações, este texto tem por objetivo construir modelos que permitam eliciar a distribuição a priori dos coeficientes de regressão discutindo razões de chance concomitante a pre-

sença de flexibilidade suficiente para lidar com as diversas discrepâncias, portanto, visa apresentar uma abordagem geral para as mais diversas situações encontradas no contexto de dados binários.

Para tanto, a perspectiva apresentada pela inferência bayesiana não-paramétrica é bastante apropriada ao considerar a distribuição que define a função de ligação como uma quantidade aleatória tal que a distribuição esperada a priori seja *Logística* e, desta forma, a elicitação a priori possa ser feita através da discussão de razões de chance. Em particular, o termo curva de probabilidade é utilizado como sinônimo para a função de distribuição que define a função de ligação ao longo deste texto.

Sob tal abordagem, é possível calcular a distribuição a posteriori da curva de probabilidade e, por isso, a função de ligação a posteriori será mais adequada do que a função *Logito*, deste modo, descrevendo apropriadamente os dados.

Neste trabalho, três modelos semi-paramétricos são apresentados. O primeiro modelo é uma mistura em escala que permite lidar com discrepâncias relacionadas a curtose do modelo *Logístico*. É uma extensão relevante a partir do que já foi proposto por [Basu e Mukhopadhyay \[1998\]](#) ao possibilitar a interpretação da distribuição a priori dos parâmetros através de razões de chances.

O segundo modelo usufrui da mistura em escala em conjunto com a transformação proposta por [Yeo e Johnson \[2000\]](#) possibilitando que a curtose assim como a assimetria sejam ajustadas e um parâmetro informativo de assimetria seja estimado. Esta transformação é muito mais apropriada para lidar com valores negativos do que a transformação de [Box e Cox \[1964\]](#) utilizada por [Guerrero e Johnson \[1982\]](#) e é mais simples do que o modelo proposto por [Stukel \[1988\]](#).

Por fim, o terceiro modelo é o mais geral entre todos e consiste em uma mistura de posição e escala tal que possa descrever curtose, assimetria e também bimodalidade. O modelo proposto por [Newton et al. \[1996\]](#), embora, seja bastante geral, não permite uma interpretação palpável da distribuição a priori para os pesquisadores da área aplicada.

E para avaliar tais modelos, três medidas de distância entre distribuições de probabilidade são sugeridas para quantificar a qualidade de ajuste dos modelos para dados binários, a saber: Cramér-von Mises, Kolmogorov-Smirnov e Anderson-Darling. Estas medidas são interessantes para analisar cada modelo sob a perspectiva de estimação de uma curva de probabilidade com o objetivo de realizar inferência. Para o objetivo de predição, as Ordenadas Preditivas Condicionais que correspondem a validação cruzada no contexto bayesiano são utilizadas.

Na sequência, o Capítulo 2 apresenta uma breve revisão de processos *Dirichlet* e transformações potência seguido pelas três sugestões de modelagem; no Capítulo 3, exemplos simulados para ilustrar o desempenho dos modelos nas situações com presença de simetria, assimetria e bimodalidade são considerados; enquanto no Capítulo 5, há duas aplicações populares na literatura como conjunto de dados desafiadores na modelagem de dados binários; o Capítulo 4 apresenta o algoritmo para o modelo semi-paramétrico com mistura em escala baseado na introdução de variáveis latentes e uma breve discussão sobre as dificuldades encontradas na utilização de tal artifício no contexto bayesiano não-paramétrico e, por fim, o Capítulo 6 apresenta considerações finais.



## Capítulo 2

# Modelos de Regressão Binária

Este capítulo tem por objetivo apresentar a modelagem de dados binários através de curvas de probabilidade em uma perspectiva bayesiana semi-paramétrica.

Para tanto, considere  $Y_{il}|p_i \sim B(p_i)$  como variáveis aleatórias binárias independentes sendo  $p_i$  definido por,

$$p_i = F(\mathbf{x}'_i \boldsymbol{\beta} | \theta, \tau), \quad (2.1)$$

para  $i = 1, \dots, n$  e  $l = 1, \dots, n_i$  tal que  $F$  é uma função de distribuição pertencente a uma família de posição e escala  $\mathbb{F} = \{F(\cdot | \theta, \tau) : \theta \in \mathfrak{R}, \tau > 0\}$ ,  $\mathbf{x}'_i$  é um vetor  $p \times 1$  de covariáveis referente ao indivíduo  $i$  e  $\boldsymbol{\beta}$  é um vetor  $p \times 1$  de parâmetros desconhecidos.

Na nomenclatura de Modelos Lineares Generalizados, a função de distribuição  $F$  é dita ser a inversa da função de ligação. Usualmente,  $F$  é simétrica com média  $\mu = 0$  e precisão  $\tau = 1$ , por exemplo, o modelo Logístico popularizado por Cox [1969] surge como consequência da suposição de que a distribuição  $F$  é a função de distribuição *Logística*(0,1),

$$p_i = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})},$$

enquanto, o modelo *Probit* introduzido por Bliss [1935] segue da suposição de que a distribuição  $F$  é a função de distribuição *Normal*(0,1),

$$p_i = \Phi(\mathbf{x}'_i \boldsymbol{\beta}).$$

Para simplificar a notação, a função de distribuição  $F(\cdot | \theta = a, \tau = b)$  será apresentada apenas como  $F(\cdot | a, b)$  para quaisquer valores de  $a$  e  $b$ . Além disso, note que  $F$  define uma curva de probabilidade ou um vetor de probabilidade de sucesso  $\mathbf{p}$  através da equação (2.1).

Sob a abordagem frequentista, a função de verossimilhança é construída e deve ser maximizada em  $\boldsymbol{\beta}$ ,

$$\begin{aligned} L(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}) &= \prod_{i=1}^n \prod_{l=1}^{n_i} p_i^{y_{il}} (1 - p_i)^{(1-y_{il})} \\ &= \prod_{i=1}^n \prod_{l=1}^{n_i} F(\mathbf{x}'_i \boldsymbol{\beta} | 0, 1)^{y_{il}} (1 - F(\mathbf{x}'_i \boldsymbol{\beta} | 0, 1))^{(1-y_{il})}, \end{aligned}$$

com  $\mathbf{Y}$  sendo um vetor  $n \times 1$  de variáveis resposta e  $\mathbf{X}$ , uma matriz de covariáveis de dimensão  $n \times p$ .

Na abordagem bayesiana paramétrica introduzida por Albert e Chib [1993], assume-se que a variável aleatória binária  $Y$  consiste na discretização de uma variável aleatória latente  $W$  e como o vetor de parâmetros  $\boldsymbol{\beta}$  é aleatório, é necessário atribuir uma distribuição a priori  $\pi(\boldsymbol{\beta})$  para que

seja possível calcular a distribuição a posteriori,

$$\pi(\boldsymbol{\beta}, \mathbf{W} | \mathbf{Y}, \mathbf{X}) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^n \prod_{l=1}^{n_i} (\mathbb{1}_{(W_{il} > 0)} \mathbb{1}_{(Y_{il}=1)} + \mathbb{1}_{(W_{il} < 0)} \mathbb{1}_{(Y_{il}=0)}) f(W_{il} | \mathbf{x}'_i \boldsymbol{\beta}, 1).$$

em que  $f(\cdot | a, b)$  é a função densidade derivada de  $F(\cdot | a, b)$ .

Como a distribuição a posteriori  $\pi(\boldsymbol{\beta}, \mathbf{W} | \mathbf{Y}, \mathbf{X})$  não apresenta uma forma fechada, são necessários métodos numéricos MCMC (Monte Carlo via Cadeia de Markov). Veja, por exemplo, [Albert e Chib \[1993\]](#), [Gamerman \[1997\]](#) e [Chen e Dey \[1998\]](#).

Em ambas abordagens, é necessário definir a função de ligação e, desta forma, diversos modelos são considerados e comparados através de critérios clássicos como AIC e BIC [[Burnham e Anderson, 2002](#)] e critérios bayesianos como CPO [[Gelfand, 1996](#)] e DIC [[Spiegelhalter et al., 2002](#)].

A abordagem bayesiana semi-paramétrica surge como um caminho alternativo ao considerar a função de ligação desconhecida, ou seja, a função de distribuição  $F$  é uma quantidade aleatória tal que se deve atribuir-lhe uma distribuição a priori.

Para tanto, a próxima seção discutirá uma classe de distribuições a priori que podem ser atribuídas a funções de distribuição.

## 2.1 Processos de Dirichlet

Considere um espaço mensurável  $(\mathbb{S}, \mathbb{A}_{\mathbb{S}})$  e uma medida de probabilidade  $H$  desconhecida sobre o mesmo. Na perspectiva bayesiana não-paramétrica,  $H$  é dita ser uma medida de probabilidade aleatória que assume valores no espaço mensurável  $(\mathbb{H}, \mathbb{A}_{\mathbb{H}})$  e, desta forma, é necessário atribuir uma distribuição a priori  $P$  para  $H$ .

[Schlaifer e Raffa \[1961\]](#) afirmam que é ideal que a classe  $\mathbb{P}$  de distribuições a priori apresente três propriedades, devendo ser, portanto:

(I1) analiticamente tratável em relação a três aspectos:

- (a) facilidade na determinação da distribuição a posteriori;
- (b) facilidade na expressão da esperança para funções de perda simples;
- (c) fechada, isto é, se a distribuição a priori é membro da classe  $\mathbb{P}$ , então, a distribuição a posteriori também é membro da mesma classe.

(I2) "rica" no sentido de que é possível expressar qualquer informação a priori.

(I3) parametrizada de tal forma que seja possível interpretar os parâmetros em relação à informação a priori.

Neste sentido, [Ferguson \[1973\]](#) introduziu o processo de Dirichlet: Seja  $\alpha H_0$  uma medida finita não-nula sobre  $(\mathbb{S}, \mathbb{A}_{\mathbb{S}})$ . É dito que  $P(\alpha, H_0)$  é um processo de Dirichlet sobre  $(\mathbb{H}, \mathbb{A}_{\mathbb{H}})$  com parâmetros  $(\alpha, H_0)$  se para todo  $k = 1, 2, \dots$ , e partição mensurável  $(A_1, \dots, A_k)$  de  $\mathbb{A}_{\mathbb{S}}$  tem-se que  $(H(A_1), \dots, H(A_k)) \sim \text{Dirichlet}(\alpha H_0(A_1), \dots, \alpha H_0(A_k))$ .

E as principais propriedades do processo de Dirichlet são,

(D1)  $P(\alpha, H_0)$  é uma medida de probabilidade sobre  $(\mathbb{H}, \mathbb{A}_{\mathbb{H}})$ ;

(D2)  $P(\alpha, H_0)$  atribui probabilidade igual a um para o conjunto de todas as medidas de probabilidade discretas sobre  $(\mathbb{S}, \mathbb{A}_{\mathbb{S}})$ ;

(D3)  $E(P(A)) = H_0(A)$  para  $A \in \mathbb{A}_{\mathbb{S}}$ ;

(D4) a distribuição a posteriori,  $P(\alpha, H_0) | S$ , é ainda um processo de Dirichlet  $P(\alpha + n, H_0 + \delta_S)$  tal que  $\delta_S$  é a medida de probabilidade degenerada em  $S$ .

Portanto, o processo de Dirichlet pode ser utilizado como uma distribuição a priori pela propriedade (D1). A propriedade (D4) satisfaz as condições desejáveis (I1.a) e (I1.c), a propriedade (D3) permite que a condição desejável (I1.b) seja satisfeita para a perda quadrática e as propriedades (D3) e (D4) satisfazem a condição desejável (I3), pois,

$$E(P(A)|S_1, \dots, S_n) = \frac{\alpha}{\alpha + n} H_0(A) + \frac{n}{\alpha + n} H_n(A) \quad \text{para } A \in \mathbb{A}_S, \quad (2.2)$$

tal que  $H_n$  é a distribuição empírica de  $S_1, \dots, S_n \stackrel{i.i.d.}{\sim} H$ . Desta forma, o parâmetro  $\alpha$  pode ser interpretado como o grau de crença na distribuição a priori, enquanto  $H_0$  é a distribuição esperada do processo de Dirichlet.

Entretanto, a propriedade (D2) parece ser inconsistente com a condição desejável (I2) quando a expressão da informação a priori ocorre através de distribuições contínuas.

Frente a tal limitação, [Antoniak \[1974\]](#) define uma mistura de processos Dirichlet que satisfaz mais convenientemente a propriedade (I2),

$$\int P(\alpha(u), H_0(A|u)) dG(u) \quad \text{para } A \in \mathbb{A}_S,$$

com  $G$  sendo uma distribuição de probabilidade. Note que a mistura de processos de Dirichlet contorna a limitação frente a distribuições contínuas, contudo, a simplicidade é perdida.

[Blackwell e MacQueen \[1973\]](#) e [Sethuraman \[1994\]](#) apresentam outras duas representações para o processo de Dirichlet. [Blackwell e MacQueen \[1973\]](#) demonstram a relação um a um entre sequências de variáveis aleatórias permutáveis e medidas de probabilidade sobre  $(\mathbb{H}, \mathbb{A}_{\mathbb{H}})$ . Em particular, o caso de variáveis aleatórias permutáveis denominado Urna de Pólia Generalizada.

Neste contexto, o processo de Dirichlet pode ser interpretado como um modelo de urnas no qual há, inicialmente,  $\alpha$  bolas pretas. Se uma bola preta é retirada, esta bola é devolvida à urna e uma nova bola de uma cor diferente é gerada a partir da distribuição uniforme e tal bola gerada é considerada como a bola retirada e adicionada à urna. Se uma bola de outra cor é retirada, esta bola não preta é devolvida à urna e uma nova bola de mesma cor é colocada na urna.

Este modelo permite estabelecer a distribuição condicional completa para qualquer variável aleatória  $S_i$  desde que a distribuição condicional completa de uma específica variável aleatória  $S$  seja conhecida. Convenientemente, para estabelecer uma equação de predição, escolhe-se  $i = n + 1$ ,

$$P(S_{(n+1)} \in A | S_1, \dots, S_n) = \frac{\alpha}{\alpha + n} H_0(A) + \frac{n}{\alpha + n} H_n(A) \quad \text{para } A \in \mathbb{A}_S, \quad (2.3)$$

Note que a equação (2.3) é idêntica à equação (2.2) e, a partir delas, é possível gerar computacionalmente o processo de Dirichlet a posteriori.

Por sua vez, [Sethuraman \[1994\]](#) apresenta uma definição construtiva do processo de Dirichlet,

$$P(A) = \sum_{k=1}^{\infty} q_k \delta_{Z_k}(A) \quad \text{para } A \in \mathbb{A}_S, \quad (2.4)$$

em que  $\delta_{Z_k}$  são medidas de probabilidade discreta concentradas em  $Z_k$  sendo que  $Z_k \sim H_0$  e os pesos  $q_k$  são variáveis aleatórias independentes de  $Z_k$  tal que

$$\begin{aligned} q_1 &= V_1, \\ q_k &= (1 - V_1)(1 - V_2) \times \dots \times (1 - V_{k-1})V_k \quad \text{para } k \geq 2, \\ \sum_{k=1}^{\infty} q_k &= 1, \end{aligned} \quad (2.5)$$

e, por sua vez,  $V_k$  são variáveis aleatórias independentes e identicamente distribuídas segundo uma distribuição  $Beta(1, \alpha)$ .

Esta construção denominada *stick-breaking* pode ser interpretada como o processo de quebra de uma vareta, inicialmente, de comprimento igual a um. Em cada etapa, a vareta é quebrada em duas partes em um ponto escolhido aleatoriamente e independentemente, de tal forma que o comprimento de uma das partes é atribuído ao peso  $q_k$  e a outra parte é novamente quebrada. O processo repete-se infinitamente.

Posteriormente, [Ishwaran e Zarepour \[2000\]](#) discutem esta construção para processos de Dirichlet finitos, isto é,

$$P(A) = \sum_{k=1}^N q_k \delta_{Z_k}(A) \quad \text{para } A \in \mathbb{A}_{\mathbb{S}}, \quad (2.6)$$

com  $N < \infty$  e fixando  $V_N = 1$  para garantir que  $\sum_{k=1}^N q_k = 1$ , pois,

$$1 - \sum_{k=1}^N q_k = (1 - V_1)(1 - V_2) \times \dots \times (1 - V_{k-1}).$$

[Ishwaran e James \[2001\]](#) definem uma classe de distribuições a priori *stick-breaking* em que  $V_k$  são variáveis aleatórias independentes e identicamente distribuídas segundo uma distribuição  $Beta(a_k, b_k)$ . Esta classe engloba diversas distribuições a priori para medidas de probabilidade aleatórias como, por exemplo, o processo com dois parâmetros Poisson-Dirichlet de [Pitman e Yor \[1997\]](#) além do processo de Dirichlet e o processo de Dirichlet finito.

Por fim, eles demonstram que para um processo construído da forma (2.4) - (2.5) ser considerado pertencente à classe de prioris *stick-breaking* é necessário que o vetor de pesos  $\mathbf{q} = (q_1, \dots, q_N)$  com  $N \leq \infty$  apresente a distribuição *Dirichlet Generalizada*  $_N(\mathbf{a}, \mathbf{b})$  discutida por [Connor e Mosimann \[1969\]](#) em que  $\mathbf{a} = (a_1, \dots, a_N)$  e  $\mathbf{b} = (b_1, \dots, b_N)$  com densidade dada por,

$$\begin{aligned} & \left( \prod_{k=1}^{N-1} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \right) q_1^{a_1-1} \times \dots \times q_{N-1}^{a_{N-1}-1} q_N^{b_{N-1}-1} \times \\ & \times (1 - Q_1)^{b_1 - (a_2 + b_2)} \times \dots \times (1 - Q_{N-2})^{b_{N-2} - (a_{N-1} + b_{N-1})} \end{aligned}$$

tal que  $Q_k = q_1 + \dots + q_k$ .

A partir de tais conceitos, pode-se apresentar o modelo proposto para curvas de probabilidade simétricas no contexto Bayesiano semi-paramétrico.

## 2.2 O primeiro modelo: mistura em escala

Como na seção anterior, considere  $Y_{il}$  uma variável aleatória binária tal que  $Y_{il}|p_i \sim B(p_i)$  sendo,

$$p_i = F(\mathbf{x}'_i \boldsymbol{\beta} | 0, \tau),$$

para  $i = 1, \dots, L$  e  $l = 1, \dots, n_i$  com  $\tau$  conhecido. A função de distribuição  $F$  é dita ser pertencente a uma família de escala  $\mathbb{F}$ .

Por simplicidade, comumente assume-se  $\tau = 1$ , contudo, é possível atribuir maior flexibilidade à curtose da função de distribuição  $F$  ao considerar  $\tau \sim H$ , ou seja, reescrever a função de distribuição  $F$  como uma mistura em escala de distribuições  $G$  simétricas e pertencentes a uma família de escala  $\mathbb{G}$  pela função de distribuição  $H$ .

Então, a função de densidade derivada da função de distribuição  $F$  é dada por,

$$f(\cdot | \mathbf{x}' \boldsymbol{\beta}, H) = \int g(\cdot | \mathbf{x}' \boldsymbol{\beta}, \tau) dH(\tau). \quad (2.7)$$

Note que  $F$  ainda é uma função de distribuição simétrica e pertencente a uma família de escala  $\mathbb{F}$ , enquanto  $H$  é uma função de distribuição não necessariamente conhecida no contexto bayesiano não-paramétrico.

Consequentemente, a função de verossimilhança é definida por,

$$L(\boldsymbol{\beta}, \boldsymbol{\tau}, H | \mathbf{Y}, \mathbf{X}) = \prod_{i=1}^L \prod_{l=1}^{n_i} p_i^{y_{il}} (1 - p_i)^{(1-y_{il})}$$

sendo que  $p_i = F(\mathbf{x}'\boldsymbol{\beta}|0, \tau)$  para  $\tau \sim H$ .

E sob a abordagem Bayesiana semi-paramétrica, distribuições a priori são atribuídas para as quantidades desconhecidas,

$$\begin{aligned} \boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \mathbf{T}_\beta &\sim N_p(\boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ H | \alpha &\sim P(\alpha, H_0). \end{aligned} \tag{2.8}$$

sendo  $H_0$  a distribuição base do processo de *Dirichlet* que é definida a partir da escolha da distribuição  $F_0$  que, por sua vez, representa a distribuição esperada da mistura em escala de distribuições  $G(\cdot|0, \tau)$  com  $\tau \sim H_0$ .

Desta forma, é essencial discutir misturas de distribuições em escala que estão bem estudadas, principalmente, quando as componentes da mistura são dadas pela distribuição *Normal*, ou seja, é razoável escolher  $G(\cdot|0, \tau)$  como a função de distribuição *Normal*(0,  $\tau$ ).

Nesta direção, segue-se [Andrews e Mallows \[1974\]](#) para escolher  $H_0$  e obter a distribuição  $F_0$  desejada. Dentre as possíveis escolhas, estão as distribuições *Gama*, *Kolmogorov-Smirnov* e *Exponencial* que resultam nas distribuições *t-student*, *Logística* e *Laplace*, respectivamente.

A escolha de  $\alpha$  pode ser feita a partir da ideia de que  $\alpha$  representa a concentração do processo de Dirichlet em torno de  $F_0$  em relação ao tamanho da amostra  $n$ , isto é, a magnitude de  $\alpha$  em relação a  $n$  define a força da crença de que  $F_0$  seja a verdadeira função de ligação.

Note que a escolha da distribuição base  $H_0$  ainda está em aberto. Neste contexto em que processos de Dirichlet são utilizados como distribuição a priori para a distribuição  $H$  na mistura em escala de distribuições dada na equação (2.7). [Basu e Mukhopadhyay \[1998\]](#) limitam-se à distribuição *Gama* devido a facilidade em obter a distribuição a posteriori, fruto da conjugação com a distribuição *Normal* em conjunto com a utilização do algoritmo Amostrador de Gibbs por Urna de Pólia através de variáveis latentes.

Esta mesma possibilidade é considerada, mas outras duas são adicionadas: as distribuições *Kolmogorov-Smirnov* e *Exponencial*.

### 2.2.1 Gama

Para  $\tau(u) = u$  sendo  $U \sim Gama(v/2, v/2)$  com densidade,

$$h_0(u) = \frac{(v/2)^{\frac{v}{2}}}{\Gamma(v/2)} Z_\tau^{\frac{v}{2}-1} \exp\left\{-\frac{v}{2}u\right\} \mathbb{1}_{(u>0)},$$

segue que  $F_0$  é dada pela distribuição  $t_v(0, 1)$ . Se  $v = 1$ ,  $F_0$  é a distribuição *Cauchy*(0, 1), enquanto se  $v \rightarrow \infty$ ,  $F_0$  é a distribuição *Normal*(0, 1) sendo que para  $v > 30$ , a distribuição *t-student* com  $v$  graus de liberdade já é considerada uma boa aproximação para a distribuição *Normal*.

### 2.2.2 Kolmogorov-Smirnov

Para obter a distribuição *Logística*(0, 1), deve-se considerar

$$\tau(u) = \left(\frac{1}{2u}\right)^2 \quad \text{tal que} \quad U \sim \text{Kolmogorov-Smirnov},$$

em que a distribuição de *Kolmogorov-Smirnov* é definida por,

$$h_0(u) = 8 \sum_{k=1}^{\infty} (-1)^{k+1} k^2 u \exp\{-2k^2 u^2\} \mathbb{1}_{(u>0)}. \quad (2.9)$$

A avaliação da densidade  $h_0$  que corresponde à densidade definida na equação (2.9) é somente possível aproximadamente. A escolha do índice de truncamento do somatório é discutida por [Chen e Dey \[1998\]](#) e depende da precisão numérica desejada  $\delta$  e da decomposição da densidade  $h_0$  em uma série monótona decrescente.

[Devroye \[1986\]](#), ao implementar o método de Séries Alternadas para gerar valores da distribuição de *Kolmogorov-Smirnov*, utiliza duas decomposições da forma,

$$h_0(u) = cf_d(u)(1 - a_1(u) + a_2(u) - a_3(u) + \dots),$$

sendo  $c$  uma constante,  $f_d$  uma distribuição fácil de ser gerada e  $a_n$  uma série monótona decrescente.

A primeira decomposição é dada por,

$$\begin{aligned} cf_d(u) &= 8u \exp\{-2u^2\}, \\ a_n(u) &= (n+1)^2 \exp\{-2u^2((n+1)^2 + 1)\}, \end{aligned}$$

em que  $a_n \searrow 0$  para  $u > \sqrt{1/3}$ , enquanto a segunda decomposição é dada por,

$$\begin{aligned} cf_d(u) &= \frac{\sqrt{2\pi}\pi^2}{4u^4} \exp\left\{-\frac{\pi^2}{8u^2}\right\}, \\ a_n(u) &= \begin{cases} \frac{4u^2}{\pi^2} \exp\left\{-\frac{(n^2-1)\pi^2}{8u^2}\right\} & \text{se } n \text{ ímpar,} \\ (n+1)^2 \exp\left\{-\frac{((n+1)^2-1)\pi^2}{8u^2}\right\} & \text{se } n \text{ par.} \end{cases} \end{aligned}$$

com  $a_n \searrow 0$  para  $u < \pi/2$ . Note que o intervalo de convergência entre tais séries apresenta uma intersecção e, seguindo [Devroye \[1986\]](#), escolhe-se  $u = 0,75$  como valor divisor para avaliar tais decomposições.

Finalmente, o índice para truncamento da densidade na equação (2.9) segue,

$$n^* = \inf\{n : cf_d(u)a_n(u) < \delta\}. \quad (2.10)$$

### 2.2.3 Exponencial

Para obter a distribuição *Laplace*, é necessário definir

$$\tau(u) = \frac{1}{2u} \quad \text{tal que } U \sim \text{Exp}(1).$$

E a primeira proposta para lidar com curvas de probabilidade assimétricas é através de funções  $\psi(\cdot; \lambda)$  pertencentes à família de transformações potência em que  $\lambda$  quantifica a assimetria presente nos dados. Desta forma, uma pequena revisão de tais famílias é apresentada, seguindo e complementando [Sakia \[1992\]](#).

## 2.3 Transformações Potência

A utilização de funções da família de transformações potência para lidar com dados assimétricos surgiu no contexto de dados quantitativos com [Tukey \[1957\]](#) ao considerar uma função monótona

para valores positivos,

$$\psi(y, \lambda) = \begin{cases} \eta^\lambda & \text{se } \lambda \neq 0 \\ \log(\eta) & \text{se } \lambda = 0 \end{cases}$$

Box e Cox [1964] modificaram tal transformação para corrigir a descontinuidade em  $\lambda = 0$ ,

$$\psi(\eta, \lambda) = \begin{cases} (\eta^\lambda - 1)/\lambda & \text{se } \lambda \neq 0 \\ \log(\eta) & \text{se } \lambda = 0 \end{cases}$$

embora esta transformação ainda esteja limitada para valores positivos.

Para lidar com esta limitação, Manly [1976] sugeriu uma alternativa baseada na função exponencial,

$$\psi(\eta, \lambda) = \begin{cases} (\exp\{\lambda\eta\} - 1)/\lambda & \text{se } \lambda \neq 0 \\ \eta & \text{se } \lambda = 0 \end{cases}$$

John e Draper [1980] apresentaram outra alternativa denominada *transformação módulo* na forma,

$$\psi(\eta, \lambda) = \begin{cases} \text{sin}(\eta)[(|\eta| + 1)^\lambda - 1]/\lambda & \text{se } \lambda \neq 0 \\ \text{sin}(\eta)\log(|\eta| + 1) & \text{se } \lambda = 0 \end{cases}$$

Seguindo esta mesma ideia, Bickel e Doksum [1981] sugeriram uma transformação para  $\lambda > 0$ ,

$$\psi(\eta, \lambda) = (\text{sin}(\eta)|\eta|^\lambda - 1)/\lambda$$

Yeo e Johnson [2000] afirmam que esta última transformação é construída para lidar com curtose ao invés de assimetria e discutem uma nova possibilidade, tomando como ponto de partida a função apresentada por John e Draper [1980] e o conceito de assimetria relativa introduzido por Zwet [1964].

Inicialmente, Zwet [1964] define que uma distribuição  $F$  é dita ser simétrica se  $F(x_0 + x) + F(x_0 - x) = 1 \quad \forall x \in I_F$  sendo que  $I_F$  é o menor intervalo para o qual  $P(X \in I_F) = 1$ . Seja  $X \sim F$  e  $Y$  é outra variável aleatória com função de distribuição  $G$ , define-se que  $\psi(x) = G^{-1}(F(x))$  tal que  $\psi(x) \sim G$ .

Neste contexto, o conceito de assimetria relativo é definido:  $G$  é uma função de distribuição mais assimétrica à direita (esquerda) do que  $F$  se  $G^{-1}(F(\cdot))$  é uma função não-decrescente convexa (côncava).

Desta forma, segue que se  $\psi(\cdot, \lambda)$  é uma função não-decrescente convexa, então, tal transformação função lida com assimetria à esquerda, enquanto, se  $\psi(\cdot, \lambda)$  é uma função não-decrescente côncava segue que tal transformação lida com a assimetria à direita.

Ao considerar a condição de que a segunda derivada da transformação seja contínua em  $\eta = 0$ , a transformação de Yeo e Johnson [2000] é dada por,

$$\psi(\eta, \lambda) = \begin{cases} [(\eta + 1)^\lambda - 1]/\lambda & \text{se } \eta \geq 0 \text{ e } \lambda \neq 0 \\ \log(\eta + 1) & \text{se } \eta \geq 0 \text{ e } \lambda = 0 \\ -[(-\eta + 1)^{2-\lambda} - 1]/(2 - \lambda) & \text{se } \eta \leq 0 \text{ e } \lambda \neq 2 \\ -\log(-\eta + 1) & \text{se } \eta \leq 0 \text{ e } \lambda = 2 \end{cases} \quad (2.11)$$

observando-se, portanto, que  $\psi(\eta, \lambda)$  é convexa para  $\lambda > 1$  e côncava para  $\lambda < 1$ .

A aplicação de funções da família de transformações potência no contexto de dados binários é bastante conhecida desde a introdução por Aranda-Ordaz [1981], seguido por Guerrero e Johnson [1982], Czado [1994] que recorrem à transformação de Box e Cox [1964] e Stukel [1988] que se

baseia na transformação definida por [Manly \[1976\]](#).

Neste sentido, a transformação introduzida por [Yeo e Johnson \[2000\]](#) parece bastante atraente e, portanto, a primeira tentativa para lidar com a assimetria estará apoiada nesta ideia.

## 2.4 O segundo modelo: transformação potência

Novamente, considere  $Y_{il}|p_i \sim B(p_i)$  como variáveis aleatórias binárias independentes sendo  $p_i$  definido por,

$$p_i = F(\psi(\mathbf{x}'_i \boldsymbol{\beta}, \lambda)|0, \tau),$$

para  $i = 1, \dots, n$  e  $l = 1, \dots, n_i$  tal que  $F$  é uma função de distribuição pertencente a uma família de escala  $\mathbb{F} = \{F(\cdot|0, \tau) : \tau > 0\}$ ,  $\psi(\cdot, \lambda)$  é a transformação definida na equação (2.11),  $\mathbf{x}'_i$  é um vetor  $p \times 1$  de covariáveis referente ao indivíduo  $i$  e  $\boldsymbol{\beta}$  é um vetor  $p \times 1$  de parâmetros desconhecidos.

A função de densidade derivada da função de distribuição  $F$  é dada por,

$$f(\cdot|\mathbf{x}'\boldsymbol{\beta}, \tau, H, \lambda) = \int \phi(\cdot|\psi(\mathbf{x}'\boldsymbol{\beta}, \lambda), \tau)dH(\tau). \quad (2.12)$$

tal que o preditor linear é transformado na mistura em escala.

A função de verossimilhança é definida por,

$$L(\boldsymbol{\beta}, \lambda, H|\mathbf{Y}, \mathbf{X}) = \prod_{i=1}^L \prod_{l=1}^{n_i} p_i^{y_{il}} (1 - p_i)^{(1-y_{il})}$$

sendo que  $p_i = F(\psi(\mathbf{x}'\boldsymbol{\beta}, \lambda)|0, \tau)$  para  $\tau \sim H$ .

E, neste caso, pode-se definir o modelo tanto na abordagem bayesiana paramétrica como não-paramétrica. Na primeira perspectiva, basta definir as distribuições a priori,

$$\begin{aligned} \boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \mathbf{T}_\beta &\sim N_p(\boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ \lambda|l_1, l_2 &\sim Normal(l_1, l_2). \end{aligned} \quad (2.13)$$

A escolha dos hiperparâmetros para a distribuição a priori de  $\lambda$  pode ser feita considerando  $\mu_\lambda = 1$  e  $\tau_\lambda = 10$  que corresponde a  $\lambda = 1$ , ou seja, a assimetria é nula.

Na abordagem bayesiana não-paramétrica, a distribuição a priori para a função de distribuição  $H$  também é considerada,

$$H|\alpha \sim P(\alpha, H_0) \quad (2.14)$$

com as mesmas escolhas de  $H_0$  já discutidas.

Desta maneira, é possível obter os modelos paramétricos *Normal*, *t-student*, *Cauchy*, *Logístico* e *Laplace* considerando assimetria na perspectiva bayesiana paramétrica assim como também pode-se defini-los a priori para  $F$  na perspectiva bayesiana não-paramétrica.

## 2.5 O terceiro modelo: mistura de posição e escala

[Lo \[1984\]](#) demonstra que qualquer densidade, isto é, com assimetria, caudas pesadas e multi-modais podem ser aproximadas pela mistura de posição e escala.

Nesta direção, a segunda estratégia para lidar com a assimetria consiste na extensão do modelo simétrico apresentado no primeiro modelo definido em (2.7) com a adição de uma mistura em posição além da mistura em escala de distribuições *Normal*.

Se  $Y_{il}|p_i \sim B(p_i)$  são variáveis aleatórias binárias independentes sendo  $p_i$  definido por,

$$p_i = F(\mathbf{x}'_i \boldsymbol{\beta}|\theta, \tau),$$

para  $i = 1, \dots, n$  e  $l = 1, \dots, n_i$  tal que  $F$  é uma função de distribuição pertencente a uma família de posição e escala  $\mathbb{F} = \{F(\cdot|\theta, \tau) : \theta \in \mathfrak{R}, \tau > 0\}$ ,  $\mathbf{x}'_i$  é um vetor  $p \times 1$  de covariáveis referente ao indivíduo  $i$  e  $\boldsymbol{\beta}$  é um vetor  $p \times 1$  de parâmetros desconhecidos.

A função de densidade derivada da função de distribuição  $F$  é dada por,

$$f(\cdot|Y, \theta + \mathbf{x}'\boldsymbol{\beta}, \tau, H) = \int \phi(\cdot|\theta + \mathbf{x}'\boldsymbol{\beta}, \tau) dH(\theta, \tau). \quad (2.15)$$

A função de verossimilhança é definida por,

$$L(\boldsymbol{\beta}, H|\mathbf{Y}, \mathbf{X}) = \prod_{i=1}^L \prod_{l=1}^{n_i} p_i^{y_{il}} (1 - p_i)^{(1-y_{il})},$$

sendo que  $p_i = F(\mathbf{x}'\boldsymbol{\beta}|\theta, \tau)$  para  $(\theta, \tau) \sim H$ .

As distribuições a priori são atribuídas às quantidades desconhecidas,

$$\begin{aligned} \boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \mathbf{T}_\beta &\sim N_p(\boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ H|\alpha &\sim P(\alpha, H_0) \end{aligned} \quad (2.16)$$

As escolhas de  $H_0$  são distribuições bivariadas para  $(\theta, \tau)$  e podem ser definidas a partir da distribuição de  $\tau$  seguida pela distribuição condicional de  $\theta|\tau$ .

Para  $H_0(\tau)$ , as possíveis escolhas são as mesmas discutidas na seção anterior, ou seja, *Gama*, *Kolmogorov-Smirnov* e *Exponencial*, enquanto a escolha natural para  $H_0(\theta|\tau)$  é dada pela distribuição  $N(0, \tau)$ .

Ferguson [1983] demonstra que a mistura de posição e escala de distribuições *Normal* considerando  $(\theta, \tau) \sim \text{Normal-Gama}$  resulta na distribuição esperada a priori  $F_0$  dada pela *t-student* com  $v$  graus de liberdade.

Raciocínio similar pode ser feito com as distribuições *Kolmogorov-Smirnov* e *Exponencial* que implica na distribuição esperada a priori  $F_0$ , respectivamente, dada pela distribuição *Logística* e *Laplace*.

## 2.6 Elicitação da distribuição a priori

A distribuição *Logística* merece atenção especial devido à facilidade na elicitação de distribuições a priori para os coeficientes de regressão.

Para estudos retrospectivos, a razão de chance é uma medida de efeito adequada e bastante popular para avaliar o grau de associação entre duas variáveis, sendo que a variável resposta é categórica, enquanto a variável explicativa pode ser qualitativa ou quantitativa.

Por exemplo, se a variável resposta  $Y$  é binária, ou seja, pode assumir as categorias  $\{A, A^c\}$  e considerando a variável explicativa qualitativa  $X$  que assume as categorias  $\{B, B^c\}$ , então, pode-se construir a Tabela 2.1.

**Tabela 2.1:** *Estudo Retrospectivo*

	$A$	$A^c$	
$B$	$n_{AB}$	$n_{A^cB}$	$n_B$
$B^c$	$n_{AB^c}$	$n_{A^cB^c}$	$n_{B^c}$
	$n_A$	$n_{A^c}$	

Como  $n_B$  e  $n_{B^c}$  não estão definidos a priori, não é razoável calcular a razão entre  $p_{A|B}$  e  $p(A|B^c)$  que é denominada risco relativo. Contudo, pode-se avaliar a medida denominada chance,

$$chance_{A|B} = \frac{p_{A|B}}{1 - p_{A|B}}$$

Deste modo, para quantificar a associação entre  $Y$  e  $X$ , a razão de chances é uma medida bastante apropriada e é definida por,

$$RC_{Y,X} = \frac{chance_{A|B}}{chance_{A|B^c}}$$

Lê-se que a chance de  $A$  para a categoria  $B$  é  $RC_{Y,X}$  maior do que a chance de  $A$  para a categoria  $B^c$ .

No contexto de regressão, as variáveis  $Y$  e  $X$  podem ser codificados como  $\{0, 1\}$  para indicar presença e ausência de um atributo. Desta forma, para a regressão *Logística* considerando  $X$  uma variável qualitativa, segue que a *chance basal* pode ser expressa através de  $\beta_0$ ,

$$\begin{aligned} chance\ basal &= \exp\{\beta_0\} \\ \beta_0 &= \log(chance\ basal). \end{aligned} \tag{2.17}$$

E a razão de chances  $RC_{Y,X}$  pode ser expressa através de  $\beta_1$ ,

$$\begin{aligned} RC_{Y,X} &= \exp\{\beta_1\} \\ \beta_1 &= \log(RC_{Y,X}). \end{aligned} \tag{2.18}$$

Se a variável  $X$  é contínua, é necessário algumas modificações. A *chance basal* ainda pode ser expressa através de  $\beta_0$  desde que a covariável  $X$  seja reparametrizada da forma  $X' = X - X_{(1)}$ , assim a equação (2.17) ainda é válida.

A razão de chances também pode ser expressa através de  $\beta_1$  pela relação (2.18), entretanto, como  $X$  é uma variável contínua segue que a interpretação da razão de chances não é simples neste caso. A razão de chances será referente ao incremento de uma unidade da variável  $X$ , o que pode ser pouco palpável para um pesquisador.

Uma solução é a utilização de um incremento de um número maior de unidades, por exemplo,  $m$  unidades. Desta forma,

$$\begin{aligned} RC_{Y,X} &= \exp\{m\beta_1\} \\ \beta_1 &= \log(RC_{Y,X})/m \end{aligned}$$

A *chance basal* assim como a *razão de chances* são quantidades pertencentes à realidade do pesquisador de uma ciência aplicada, principalmente em um estudo retrospectivo. Uma vez que tais quantidades são expressas em função dos coeficientes de regressão, a elicitação da distribuição a priori para os parâmetros segue diretamente.

Para estudos de toxicidade, a informação a priori pode ser obtida através da probabilidade basal de toxicidade e a dose letal mediana. Neste tipo de estudo, a covariável  $X$  é sempre quantitativa.

A probabilidade basal de toxicidade pode ser reescrita como função de  $\beta_0$ , sendo necessário novamente reparametrizar a covariável  $X$  através da relação  $X' = X - X_{(1)}$ , portanto, segue que a

probabilidade basal de toxicidade é dada por

$$\begin{aligned} p_{basal} &= F(\beta_0) \\ \beta_0 &= F^{-1}(p_{basal}) \end{aligned}$$

Enquanto a dose letal mediana pode ser reescrita como função de  $\beta_1$ ,

$$\begin{aligned} 0,5 &= F(\beta_0 + \beta_1 LD50) \\ F^{-1}(0,5) &= \beta_0 + \beta_1 LD50 \\ \beta_1 &= \frac{F^{-1}(0,5) - F^{-1}(p_{basal})}{LD50} \end{aligned}$$

Note que a elicitação para os coeficientes de regressão em estudos de toxicidade é aplicável para qualquer  $F$ .

## 2.7 Distribuição a posteriori

O passo seguinte é o cálculo da distribuição a posteriori,

$$\pi(\boldsymbol{\beta}, H | \mathbf{Y}, \mathbf{X}) \propto L(\boldsymbol{\beta}, H | \mathbf{Y}, \mathbf{X}) \pi(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \mathbf{T}_\beta) P(\alpha, H_0 | v),$$

que pode ser construída a partir da amostragem das distribuições condicionais completas a posteriori,

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, H, \boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ \pi(H | \mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}). \end{aligned}$$

Note que a distribuição a posteriori de  $H$  neste contexto não é um simples processo de Dirichlet, mas uma mistura de processos Dirichlet, como demonstrado por [Antoniak \[1974\]](#).

Em contrapartida ao processo de Dirichlet, uma mistura de processos de Dirichlet não é simples de ser gerada computacionalmente e se torna necessário a utilização de algoritmos mais elaborados, por exemplo, o Amostrador de Gibbs da Urna de Pólia elaborado por [Escobar \[1994\]](#) e o Amostrador de Gibbs em Blocos por [Ishwaran e James \[2001\]](#).

Em particular, o segundo algoritmo utiliza um truncamento do processo *Dirichlet*, isto é, através de um processo finito busca aproximar um processo infinito e permite a introdução de passos de Metropolis-Hastings diferentemente do primeiro algoritmo, o que torna o algoritmo Amostrador de Gibbs por Blocos bastante interessante.

Neste capítulo, para contornar preocupações devido ao aspecto computacional, optou-se por utilizar o algoritmo Metropolis-Hastings Adaptativo com uma distribuição proposta multivariada implementado no pacote estatístico LaplacesDemon, que fornece um ambiente amigável para o desenvolvimento de inferência bayesiana ao disponibilizar diversos algoritmos MCMC programados em  $C++$  dentro do programa R (R Core Team, 2004).

Para tanto, é necessário considerar um processo *Dirichlet* finito definido na equação (2.6) e reescrever os modelos através da representação *stick-breaking* dada nas equações (2.4) e (2.5).

O modelo de mistura em escala,

$$F(\mathbf{x}'\boldsymbol{\beta}) = \sum_{k=1}^N q_k \Phi(\mathbf{x}'\boldsymbol{\beta} | 0, \tau_k), \quad (2.19)$$

O modelo de mistura em escala com transformação,

$$F(\psi(\mathbf{x}'\boldsymbol{\beta}, \lambda)) = \sum_{k=1}^N q_k \Phi(\psi(\mathbf{x}'\boldsymbol{\beta}, \lambda) | 0, \tau_k), \quad (2.20)$$

Por último, o modelo de posição e escala,

$$F(\mathbf{x}'\boldsymbol{\beta}) = \sum_{k=1}^N q_k \Phi(\mathbf{x}'\boldsymbol{\beta}|\theta_k, \tau_k). \quad (2.21)$$

## 2.8 Estimação pontual de $\boldsymbol{\beta}$ , $\lambda$ e $F$

O vetor de parâmetros  $\boldsymbol{\beta}$  pode ser estimado por,

$$\hat{\boldsymbol{\beta}}|\mathbf{Y} = \frac{1}{B} \sum_{b=1}^B \boldsymbol{\beta}^{(b)}.$$

A estimação do parâmetro da transformação no segundo modelo,

$$\hat{\lambda}|\mathbf{Y} = \frac{1}{B} \sum_{b=1}^B \lambda^{(b)}.$$

A estimação da função de distribuição  $F$  segue para o primeiro modelo definido em (2.7),

$$\hat{F}(\mathbf{x}_i'\boldsymbol{\beta})|\mathbf{Y}, \mathbf{X} = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}_i'\boldsymbol{\beta}^{(b)}|0, \tau_k^{(b)}).$$

Para o segundo modelo apresentado em (2.12),

$$\hat{F}(\psi(\mathbf{x}_i'\boldsymbol{\beta}, \lambda))|\mathbf{Y}, \mathbf{X} = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\psi(\mathbf{x}_i'\boldsymbol{\beta}^{(b)}, \lambda^{(b)})|0, \tau_k^{(b)}).$$

E para o terceiro modelo em (2.15),

$$\hat{F}(\mathbf{x}_i'\boldsymbol{\beta})|\mathbf{Y}, \mathbf{X} = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}_i'\boldsymbol{\beta}^{(b)}|\theta_k^{(b)}, \tau_k^{(b)}).$$

## 2.9 Distribuição Preditiva

A distribuição preditiva  $P(Y_{(n+1)l} = 1|\mathbf{Y}, \mathbf{X})$  pode ser aproximada para o primeiro modelo por,

$$P(Y_{(n+1)l} = 1|\mathbf{Y}, \mathbf{X}) = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}'_{(n+1)l} \boldsymbol{\beta}^{(b)}|0, \tau_k^{(b)}),$$

Para o segundo modelo,

$$P(Y_{(n+1)l} = 1|\mathbf{Y}, \mathbf{X}) = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\psi(\mathbf{x}'_{(n+1)l} \boldsymbol{\beta}^{(b)}, \lambda^{(b)})|0, \tau_k^{(b)}).$$

E para o terceiro modelo,

$$P(Y_{(n+1)l} = 1|\mathbf{Y}, \mathbf{X}) = \frac{1}{B} \sum_{b=1}^B \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}'_{(n+1)l} \boldsymbol{\beta}^{(b)}|\theta_k^{(b)}, \tau_k^{(b)}).$$

## 2.10 Ordenada Preditiva Condicional

O cálculo da  $i$ -ésima Ordenada Preditiva Condicional é baseado sobre  $S_i = \sum_{l=1}^{n_i} Y_{il}$  que apresenta distribuição *Binomial*( $n_i, p_i$ ) tal que  $S_1, \dots, S_L$  são variáveis aleatórias independentes e ditas serem estatísticas suficientes para  $p_1, \dots, p_L$  sendo que  $p_i = F(\mathbf{x}'_i \beta | \beta)$ .

As Ordenadas Preditivas Condicionais (OPC) expressam a probabilidade de observar  $S_i$  quando o modelo é ajustado considerando todos os dados com exceção de  $S_i$ , isto é, a ideia corresponde a uma validação cruzada com a retirada de uma observação.

Se  $\mathbf{S}_{[i]}$  é o vetor das variáveis  $S_1, \dots, S_L$  retirando  $S_i$ , então, a  $i$ -ésima OPC é definida por,

$$P(S_i = s_i | \mathbf{S}_{[i]}) = \binom{n_i}{s_i} (p_i)^{s_i} (1 - p_i)^{n_i - s_i},$$

com

$$p_i = \hat{F}(\mathbf{x}'_i \hat{\beta}) | \mathbf{S}_{[i]}, \mathbf{X}_{[i]}.$$

Entretanto tal estratégia apresenta um alto custo computacional, ao exigir que o processo de estimação seja repetido  $L$  vezes tal que em cada repetição um conjunto de observações seja considerado.

Para evitar tal esforço, pode-se calcular a  $i$ -ésima ordenada preditiva através da média harmônica da função de verossimilhança,

$$P(S_i = s_i | \mathbf{S}_{[i]}) = \frac{1}{\frac{1}{b} \sum_{b=1}^B [P(S_i = s_i | p_i^{(b)})]^{-1}},$$

com,

$$p_i^{(b)} = \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}'_i \beta^{(b)} | 0, \tau_k^{(b)}),$$

para o primeiro modelo.

Para o segundo,

$$p_i^{(b)} = \sum_{k=1}^N q_k^{(b)} \Phi(\psi(\mathbf{x}'_i \beta^{(b)}, \lambda^{(b)}) | 0, \tau_k^{(b)}).$$

E para o terceiro,

$$p_i^{(b)} = \sum_{k=1}^N q_k^{(b)} \Phi(\mathbf{x}'_i \beta^{(b)} | \theta_k^{(b)}, \tau_k^{(b)}).$$

Pode-se reescalonar os valores de OPC ao dividi-los pelo valor máximo e, desta forma, utilizá-los como uma medida de qualidade de ajuste, sendo que valores muito pequenos indicam que o modelo não é capaz de descrever determinada observação adequadamente. Além disso, a soma do logaritmo das OPC é um estimador para o log-verossimilhança marginal que permite construir um pseudo-fator de Bayes para que modelos não-aninhados sejam comparados.

## 2.11 Distâncias entre medidas de probabilidade

Além da avaliação dos modelos através das Ordenadas Preditivas Condicionais, é factível distinguir os modelos através de medidas de distância entre distribuições de probabilidade, uma vez

que estas são sinônimos das curvas de probabilidade. Há diversas distâncias na literatura, este texto considerou as medidas de Cramér-von Mises, Anderson-Darling, Kolmogorov-Smirnov.

Neste sentido, seja  $P$  e  $Q$  duas medidas de probabilidade com funções de densidade dadas por  $p$  e  $q$  tal que  $Q$  é a curva de probabilidade empírica, enquanto  $P$  é a distribuição proposta como geradora de  $Q$ .

### Cramér-von Mises

É a medida de distância mais simples entre distribuições de probabilidades introduzida por Cramer [1928] e von Mises [1928].

$$D_{CM}(P, Q) = \sum_{i=1}^n p(x_i)[Q(x_i) - P(x_i)]^2.$$

### Kolmogorov-Smirnov

Kolmogorov [1933] e Smirnov [1948] apresentaram uma medida alternativa,

$$D_{KS}(P, Q) = \sup_x |P(x) - Q(x)|.$$

### Anderson-Darling

Anderson e Darling [1952] introduziram uma medida similar à distância de Cramér-von Mises, contudo, há um maior peso nas observações pertencentes à cauda da distribuição,

$$D_{AD}(P, Q) = \sum_{i=1}^n p(x_i) \frac{[Q(x_i) - P(x_i)]^2}{P(x_i)[1 - P(x_i)]}.$$

## Capítulo 3

# Exemplos Simulados

Três exemplos simulados para ilustrar os modelos propostos são apresentados neste capítulo. Para todos os exemplos, uma amostra  $Y_i \sim \text{Binomial}(n, p_i)$  foi utilizada para  $n = 10, 30$  e  $p_i$  dado por (3.1), (3.2) e (3.3) considerando uma única covariável  $x_i \sim N(0, 1)$  para  $i = 1, \dots, 20$ .

O primeiro exemplo simulado é dado pela distribuição *Logística*, que corresponde a uma curva de probabilidade simétrica,

$$F(\mathbf{x}\boldsymbol{\beta}) = \frac{\exp\{\beta_0 + \beta_1 x_i\}}{1 + \exp\{\beta_0 + \beta_1 x_i\}}. \quad (3.1)$$

O segundo exemplo simulado é definido pela distribuição *Gumbel* para o máximo que é uma curva de probabilidade assimétrica,

$$F(\mathbf{x}\boldsymbol{\beta}) = \exp\{-\exp\{-(\beta_0 + \beta_1 x_i)\}\} \quad (3.2)$$

sendo que  $\beta_0 = 1$  e  $\beta_1 = 1.5$  para ambos os modelos.

O último exemplo simulado consiste na mistura de distribuições *Logística* que retrata uma curva de probabilidade bimodal,

$$F(\mathbf{x}\boldsymbol{\beta}) = w \frac{\exp\{\beta_{01} + \beta_{11} x_i\}}{1 + \exp\{\beta_{01} + \beta_{11} x_i\}} + (1 - w) \frac{\exp\{\beta_{02} + \beta_{12} x_i\}}{1 + \exp\{\beta_{02} + \beta_{12} x_i\}} \quad (3.3)$$

tal que  $\beta_{01} = -6$ ,  $\beta_{11} = 5$ ,  $\beta_{02} = 6$ ,  $\beta_{12} = 5$  e  $w = 0,5$ .

O pacote *LaplacesDemon* permitiu o ajuste dos modelos paramétricos e semi-paramétricos pelo algoritmo Metropolis-Hastings Adaptativo tal que uma cadeia de 10000 iterações foi utilizada na fase de adaptação com a estimação da matriz de covariâncias a cada 1000 iterações.

O tamanho da amostra de aquecimento, assim como o espaçamento entre as amostras coletadas nas cadeias de Markov foram definidos para que houvesse convergência através da análise gráfica e a autocorrelação fosse próxima a zero.

A distribuição a priori para  $\beta_0$  e  $\beta_1$  consideradas foram independentes e não-informativas ao defini-las como *Normal* com média 0 e precisão 0,01. Se houvesse informação a priori, a elicitação poderia ser feita conforme discutido no Capítulo 2.

Para os modelos que recorrem à utilização da transformação, a distribuição a priori de  $\lambda$  é definida como a distribuição *Normal*(1, 0, 01), desta forma, é esperado uma curva de probabilidade simétrica a priori e acredita-se tanto na assimetria à esquerda como à direita. Note que a informação sobre  $\lambda$  poderia ser elicitada a partir da discussão da forma da curva de probabilidade.

Para o modelo de mistura em escala,  $H_0(\tau)$  corresponde à distribuição de Kolmogorov-Smirnov e para o modelo de mistura em posição e escala,  $H_0(\theta|\tau)$  é definida pela distribuição *Normal*(0,  $\tau$ ), isto é, a curva de probabilidade esperada a priori é definida pela distribuição *Logística*. Para todos os modelos semiparamétricos, foi estabelecido  $\alpha = 1$ , portanto, a crença de que a distribuição *Logística* seja apropriada aos dados é bastante fraca.

Os modelos bayesianos *Logístico* paramétrico (BP), paramétrico com transformação (BP TF),

semi-paramétrico com mistura em escala (BSP ME), semi-paramétrico com transformação (BSP TF) e semi-paramétrico com mistura em posição e escala (BSP MEP) foram ajustados para cada conjunto de dados.

### 3.1 Modelo Logístico

Este exemplo corresponde a um problema simétrico e, por isso, é esperado que todos os modelos apresentem um desempenho similar. As estimativas estão apresentadas nas Tabelas 3.1 e 3.3 para  $n = 10, 30$ , respectivamente.

É possível recuperar os valores do vetor do parâmetro  $\beta$  simulados neste exemplo, com exceção da estimativa pontual para  $\beta_0$  no modelo BSP MEP em que o intervalo de credibilidade 95% contém tanto valores positivos como negativos. Note também que a estimativa intervalar do parâmetro  $\lambda$  nos dois modelos BP TF e BSP TF contém o modelo simétrico.

Somente no modelo BP é possível interpretar os parâmetros a posteriori através do conceito de razão de chances, todavia, ainda é possível calcular razões de chances através das curvas de probabilidade estimadas. Na Figura 3.1 estão as curvas de probabilidade estimada para 30 ensaios *Bernouli*.

Além disso, todos os modelos apresentam desempenhos similares para todas as medidas de distância, assim como as ordenadas preditivas condicionais, embora possa se dar um pequeno destaque para o modelo MEP.

O aumento do número de ensaios *Bernouli* possibilita melhores estimativas da curva de probabilidade, como é verificado pelas medidas de distância nas Tabelas 3.2 e a redução das mesmas na Tabela 3.4.

**Tabela 3.1:** *Dados do modelo Logístico com 10 ensaios Bernouli - Estimativas para os modelos*

Modelo	Estimativa	IC 95%	SLOPC
<i>Logistic</i> BP			-26,339
$\beta_0$	1,125	(0,641 ; 1,638)	
$\beta_1$	1,993	(1,468 ; 2,594)	
<i>Logistic</i> BP TF			-26.607
$\beta_0$	1,380	(0,636 ; 2,404)	
$\beta_1$	2,093	(1,543 ; 2,807)	
$\lambda$	0,746	(0,222 ; 1,326)	
<i>Logistic</i> BSP ME			-26,255
$\beta_0$	0,751	(0,382 ; 1,279)	
$\beta_1$	1,365	(0,863 ; 2,219)	
<i>Logistic</i> BSP TF			-26.512
$\beta_0$	1.002	(0.411 ; 1.989)	
$\beta_1$	1.484	(0.910 ; 2.576)	
$\lambda$	0.607	(-0.094 ; 1.341)	
<i>Logistic</i> BSP MEP			-26,039
$\beta_0$	0,707	(-0,788 ; 1,846)	
$\beta_1$	1,592	(1,006 ; 2,452)	

**Tabela 3.2:** Dados do modelo Logístico com 10 ensaios Bernouli - Medidas de distância e redução percentual relativa

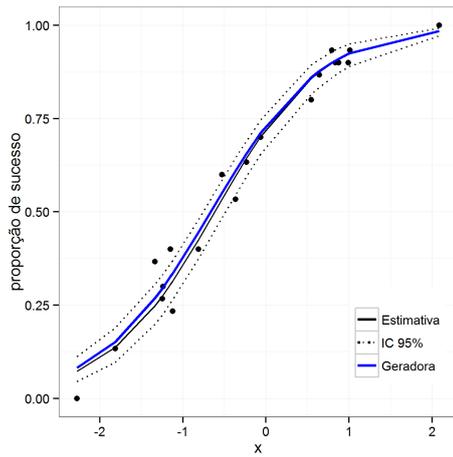
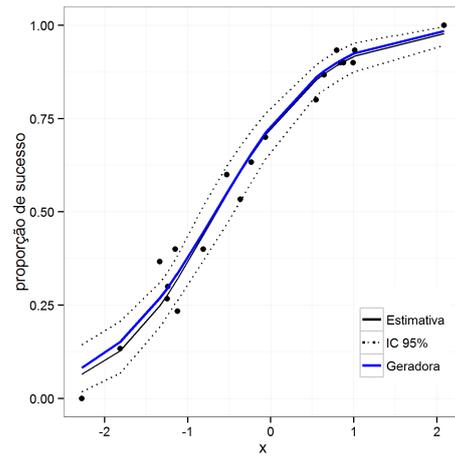
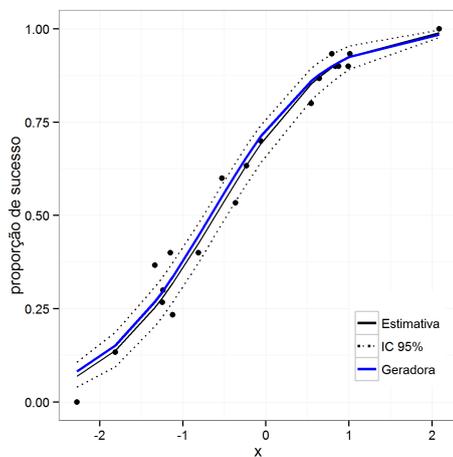
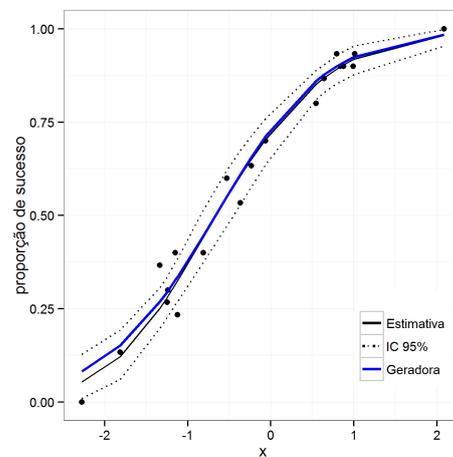
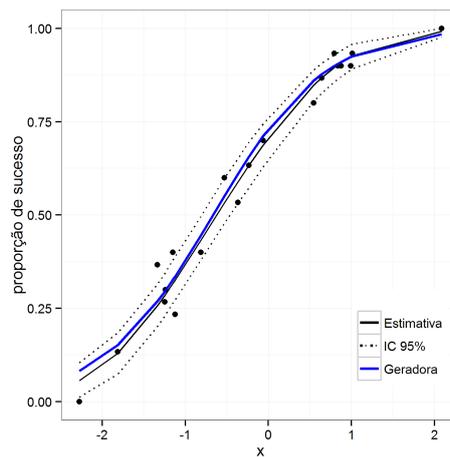
Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP 0,00745 (0,0)	0,15022 (0,0)	0,05687 (0,0)	
BP TF 0,00795 (-6,7)	0,16170 (-7,6)	0,05091 (10,5)	
BSP ME 0,00764 (-2,6)	0,15665 (-4,3)	0,05661 (0,4)	
BSP TF 0,00864 (-16,0)	0,16477 (-9,7)	0,05476 (3,7)	
BSP MEP 0,00757 (-1,6)	0,15634 (-4,1)	0,05380 (5,4)	

**Tabela 3.3:** Dados do modelo Logístico com 30 ensaios Bernouli - Estimativas para os modelos

Modelo	Estimativa	IC 95%	SLOPC
<i>Logístico</i> BP			-38,274
$\beta_0$	0,958	(0,729 ; 1,195)	
$\beta_1$	1,551	(1,300 ; 1,801)	
<i>Logístico</i> BP TF			-39,770
$\beta_0$	1,037	(0,630 ; 1,517)	
$\beta_1$	1,585	(1,280 ; 1,897)	
$\lambda$	0,911	(0,516 ; 1,368)	
<i>Logístico</i> BSP ME			-37,866
$\beta_0$	0,644	(0,415 ; 1,001)	
$\beta_1$	1,058	(0,745 ; 1,578)	
<i>Logístico</i> BSP TF			-38,939
$\beta_0$	0,729	(0,381 ; 1,249)	
$\beta_1$	1,104	(0,740 ; 1,755)	
$\lambda$	0,841	(0,348 ; 1,417)	
<i>Logístico</i> BSP MEP			-37,742
$\beta_0$	0,644	(-0,872 ; 1,913)	
$\beta_1$	1,324	(0,858 ; 2,166)	

**Tabela 3.4:** Dados do modelo Logístico com 30 ensaios Bernouli - Medidas de distância e redução percentual relativa

Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,00363 (0,0)	0,11846 (0,0)	0,02443 (0,0)
BP TF	0,00359 (1,1)	0,11875 (0,0)	0,02358 (3,4)
BSP ME	0,00338 (6,9)	0,11355 (4,4)	0,02217 (9,2)
BSP TF	0,00348 (4,1)	0,11612 (2,2)	0,02187 (10,4)
BSP MEP	0,00319 (12,1)	0,11020 (7,2)	0,01986 (18,7)

(a) *modelo BP*(b) *modelo BP TF*(c) *modelo BSP ME*(d) *modelo BSP TF*(e) *modelo BSP MEP*

**Figura 3.1:** Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo Logístico com 30 ensaios Bernouli

### 3.2 Modelo Log-log

Neste exemplo há presença de assimetria à esquerda, desta forma, é razoável que os modelos simétricos (BP e BSP ME) não gerem boas estimativas da curva de probabilidade. As estimativas estão apresentadas nas Tabelas 3.5 e 3.7 para  $n = 10, 30$ .

Não é possível recuperar os valores do vetor do parâmetro  $\beta$  simulados neste exemplo, o que era esperado, pois o modelo ajustado não é mais o mesmo modelo que gerou os dados.

A estimativa intervalar 95% do parâmetro de assimetria  $\lambda$  para ambos os modelos (BP TF e BSP TF) contém o modelo simétrico com pequena probabilidade quando se considera 10 ensaios *Bernouli*, sendo que a estimativa pontual indica assimetria à esquerda. Se 30 ensaios *Bernouli* são considerados, o intervalo de credibilidade 95% descarta o modelo simétrico. As curvas de probabilidade estimadas para esta última situação estão na Figura 3.2.

Os modelos simétricos apresentam os piores desempenhos através das medidas de distância apresentada nas Tabelas 3.6 e 3.8, isto é, maiores distância entre a distribuição definida pelos dados e a proposta pelos modelos. Dentre os modelos assimétricos, o modelo BP TF é aquele com maiores ganhos percentuais na redução da distância, seguido pelo modelo BSP TF e, por último, o modelo BSP MEP, embora as diferenças sejam bastante pequenas entre estes modelos. As Ordenadas Preditivas Condicionais possibilitam conclusões similares.

**Tabela 3.5:** *Dados do modelo Log-log com 10 ensaios Bernouli - Estimativas para os modelos*

Modelo	Estimativa	IC 95%	SLOPC
<i>Logistic</i> BP			-26,244
$\beta_0$	0,670	(0,232 ; 1,161)	
$\beta_1$	2,221	(1,711 ; 2,844)	
<i>Logistic</i> BP TF			-25.290
$\beta_0$	1.147	(0,419 ; 2,040)	
$\beta_1$	2.415	(1,775 ; 3,188)	
$\lambda$	0.551	(0,092 ; 1,019)	
<i>Logistic</i> BSP ME			-26,311
$\beta_0$	0,437	(0,147 ; 0,831)	
$\beta_1$	1,516	(1,004 ; 2,374)	
<i>Logistic</i> BSP TF			-25.652
$\beta_0$	0.773	(0,260 ; 1,564)	
$\beta_1$	1.627	(1,019 ; 2,605)	
$\lambda$	0.403	(-0,231 ; 1,004)	
<i>Logistic</i> BSP MEP			-25,758
$\beta_0$	0,707	(-0,788 ; 1,846)	
$\beta_1$	1,592	(1,006 ; 2,452)	

**Tabela 3.6:** *Dados do modelo Log-log com 10 ensaios Bernouli - Medidas de distância e redução percentual relativa*

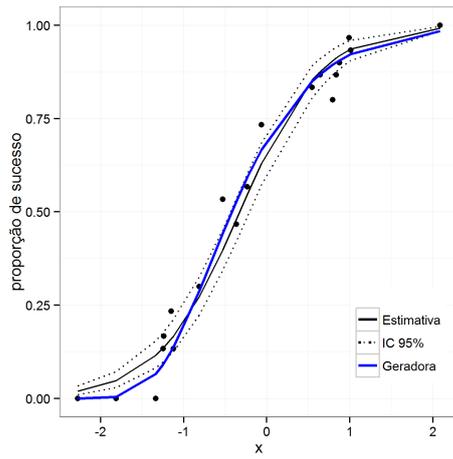
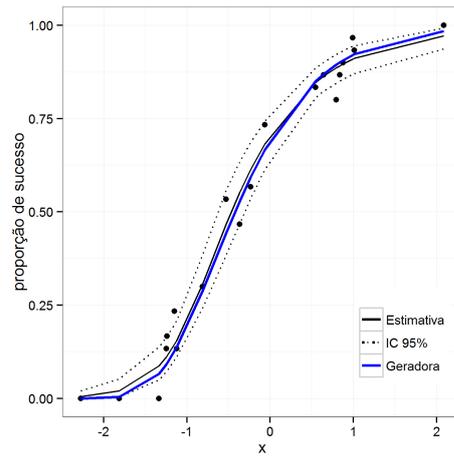
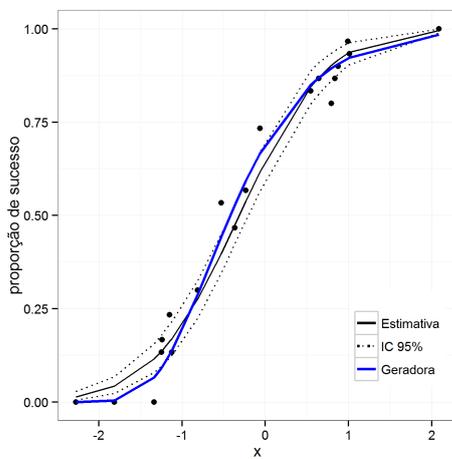
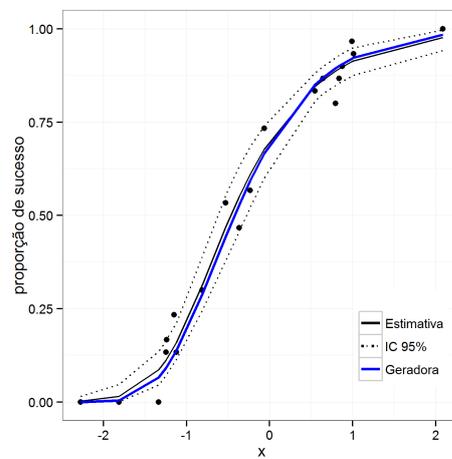
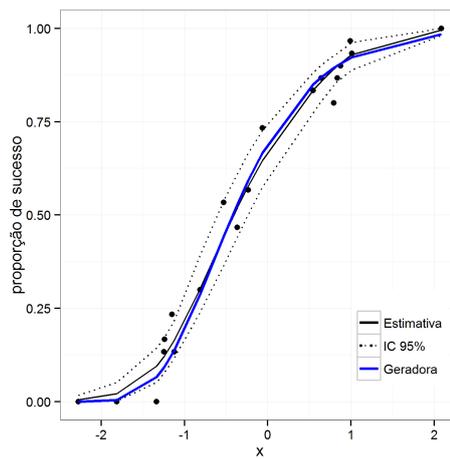
Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,00866 (0,0)	0,26340 (0,0)	0,05175 (0,0)
BP TF	0,00539 (37,6)	0,18206 (30,9)	0,04288 (17,1)
BSP ME	0,00877 (-1,3)	0,27252 (-3,5)	0,05106 (1,3)
BSP TF	0,00548 (36,7)	0,18479 (29,8)	0,04511 (12,8)
BSP MEP	0,00687 (20,1)	0,24051 (8,7)	0,04482 (13,4)

**Tabela 3.7:** *Dados do modelo Log-log com 30 ensaios Bernouli - Estimativas para os modelos*

<b>Modelo</b>	<b>Estimativa</b>	<b>IC 95%</b>	<b>SLOPC</b>
<i>Logistico BP</i>			-39,131
$\beta_0$	0,663	(0,407 ; 0,911)	
$\beta_1$	2,029	(1,753 ; 2,326)	
<i>Logistico BP TF</i>			-37.709
$\beta_0$	1,031	(0,603 ; 1,559)	
$\beta_1$	2,185	(1,835 ; 2,577)	
$\lambda$	0,636	(0,321 ; 0,965)	
<i>Logistico BSP ME</i>			-38,749
$\beta_0$	0,438	(0,242 ; 0,718)	
$\beta_1$	1,396	(0,955 ; 2,031)	
<i>Logistico BSP TF</i>			-37.007
$\beta_0$	0.707	(0.342 ; 1.305)	
$\beta_1$	1.500	(1.032 ; 2.289)	
$\lambda$	0.526	(0.089 ; 0.969)	
<i>Logistico BSP MEP</i>			-36,879
$\beta_0$	0,196	(-1,317 ; 1,730)	
$\beta_1$	1,858	(1,190 ; 2,901)	

**Tabela 3.8:** *Dados do modelo Log-log com 30 ensaios Bernouli - Medidas de distância e redução percentual relativa*

<b>Modelo</b>	<b>Cramér Von-Mises</b>	<b>Kolmogorov-Smirnov</b>	<b>Anderson-Darling</b>
BP	0,00489 (0,0)	0,13497 (0,0)	0,03114 (0,0)
BP TF	0,00272 (44,4)	0,08833 (34,6)	0,0199 (36,1)
BSP ME	0,00501 (-2,5)	0,13582 (-0,6)	0,03127 (-0,4)
BSP TF	0,00265 (45,8)	0,08692 (35,6)	0,01953 (37,3)
BSP MEP	0,00311 (37,9)	0,09513 (29,5)	0,02135 (31,4)

(a) *modelo BP*(b) *modelo BP TF*(c) *modelo BSP ME*(d) *modelo BSP TF*(e) *modelo BSP MEP*

**Figura 3.2:** Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo Log-log com 30 ensaios Bernoulli

### 3.3 Modelo de Mistura *Logístico*

Por fim, este último exemplo consiste na mistura de distribuições tal que a distribuição resultante seja bimodal. Com exceção do modelo BSP MEP, os outros modelos não são contruídos para descrever tal situação singular. As estimativas estão apresentadas nas Tabelas 3.9 e 3.11 para  $n = 10, 30$ .

Como esperado, não é possível recuperar os valores definidos na simulação. A estimativa intervalar do parâmetro  $\beta_0$  contém tanto valores positivos como negativos para todos os modelos.

Para os modelos com transformação, as estimativas pontuais do parâmetro de assimetria  $\lambda$  indicam o modelo simétrico, assim como as estimativas intervalares considerando 10 ensaios *Bernouli*. Quando se considera 30 ensaios *Bernouli*, os intervalos de credibilidade 95% para o parâmetro de assimetria descartam o modelo simétrico, indicando assimetria à esquerda.

Todos os modelos apresentam desempenho similares quando somente 10 ensaios *Bernouli* são utilizados na estimação da curva de probabilidade, como se observa na Tabela 3.10. Por outro lado, segue da Tabela 3.12 que o modelo BSP MEP apresenta notavelmente o melhor ajuste para 30 ensaios *Bernouli*, enquanto os outros modelos não são capazes de lidar com a bimodalidade da curva de probabilidade.

**Tabela 3.9:** *Dados do modelo de Mistura Logístico com 10 ensaios Bernouli - Estimativas para os modelos*

Modelo	Estimativa	IC 95%	SLOPC
<i>Logistic</i> BP			-32,947
$\beta_0$	-0,262	(-0,570 ; 0,055)	
$\beta_1$	0,989	(0,667 ; 1,307)	
<i>Logistic</i> BP TF			-35,744
$\beta_0$	-0.033	(-1.168 ; 1.076)	
$\beta_1$	0.938	(0.349 ; 1.813 )	
$\lambda$	-0.434	(-5.935 ; 2.228)	
<i>Logistic</i> BSP ME			-32,601
$\beta_0$	-0,193	(-0,456 ; 0,027)	
$\beta_1$	0,701	(0,423 ; 1,110)	
<i>Logistic</i> BSP ME TF			-35,810
$\beta_0$	-0.056	(-0.929 0.726)	
$\beta_1$	0.675	(0.242 1.395)	
$\lambda$	-0.593	(-7.329 ; 2.826)	
<i>Logistic</i> BSP MEP			-31,992
$\beta_0$	-0,205	(-2,343 ; 2,023)	
$\beta_1$	1,181	(0,509 ; 2,519)	

**Tabela 3.10:** *Dados do modelo Logístico com 10 ensaios Bernouli - Medidas de distância e redução percentual relativa*

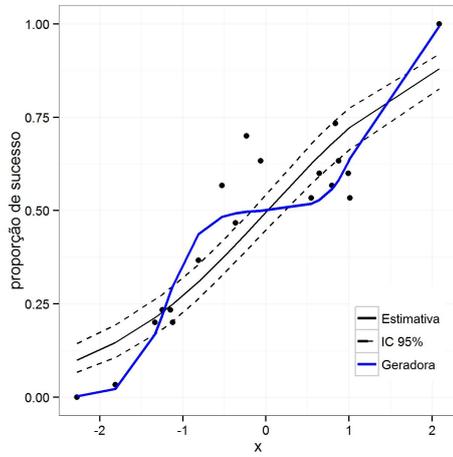
Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,01206 (0,0)	0,24199 (0,0)	0,07939 (0,0)
BP TF	0,01423 (-18,0)	0,25179 (-4,0)	0,07862 (0,9)
BSP ME	0,01199 (0,6)	0,24253 (-0,2)	0,07932 (0,1)
BSP TF	0,01403 (-16,3)	0,23418 (3,2)	0,07946 (-0,1)
BSP MEP	0,01039 (13,8)	0,21794 (9,9)	0,07216 (9,1)

**Tabela 3.11:** *Dados do modelo de Mistura Logístico com 30 ensaios Bernouli - Estimativas para os modelos*

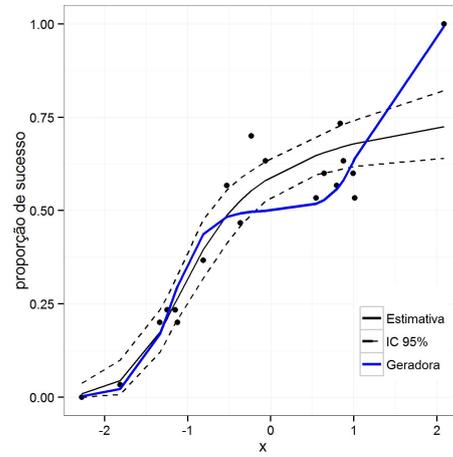
Modelo	Estimativa	IC 95%	SLOPC
<i>Logístico</i> BP			-58,984
$\beta_0$	-0,016	(-0,205 ; 0,171)	
$\beta_1$	0,970	(0,792 ; 1,146)	
<i>Logístico</i> BP TF			-53,677
$\beta_0$	0,532	(0,163 ; 1,010)	
$\beta_1$	1,069	(0,758 ; 1,427)	
$\lambda$	-0,533	(-1,460 ; 0,258)	
<i>Logístico</i> BSP ME			-57,858
$\beta_0$	-0,015	(-0,151 ; 0,117)	
$\beta_1$	0,696	(0,475 ; 1,037)	
<i>Logístico</i> BSP ME TF			-53,692
$\beta_0$	0,362	(0,078 ; 0,793)	
$\beta_1$	0,773	(0,458 ; 1,376)	
$\lambda$	-0,851	(-2,316 ; 0,251)	
<i>Logístico</i> BSP MEP			-42,201
$\beta_0$	1,275	(-2,172 ; 3,404)	
$\beta_1$	2,219	(1,460 ; 3,235)	

**Tabela 3.12:** *Dados do modelo de Mistura Logístico com 30 ensaios Bernouli - Medidas de distância e redução percentual relativa*

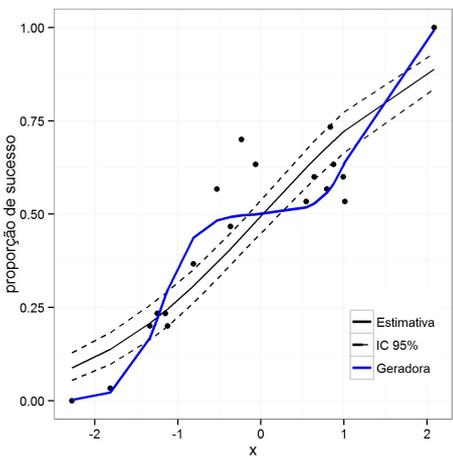
Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,01184 (0,0)	0,26049 (0,0)	0,07121 (0,0)
BP TF	0,00617 (47,8)	0,27516 (-5,6)	0,02919 (59,0)
BSP ME	0,01135 (4,1)	0,26141 (-0,4)	0,06816 (4,3)
BSP TF	0,00649 (10,5)	0,26562 (5,2)	0,03088 (14,8)
BSP MEP	0,00203 (82,8)	0,1476 (10,3)	0,01513 (78,9)



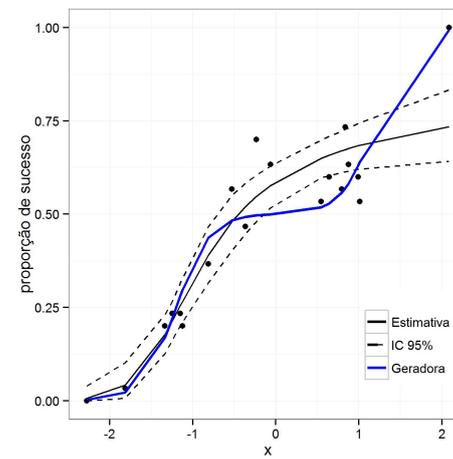
(a) modelo BP



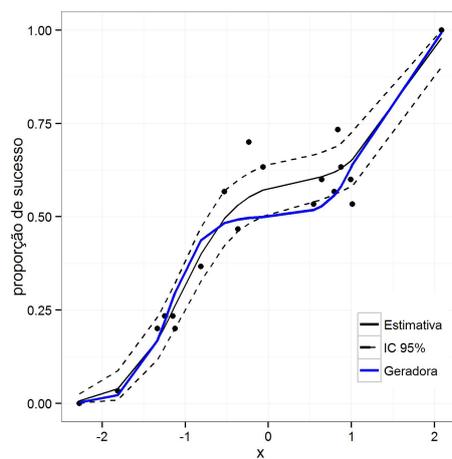
(b) modelo BP TF



(c) modelo BSP ME



(d) modelo BSP TF



(e) modelo BSP MEP

**Figura 3.3:** Estimativas das curvas de probabilidade do conjunto de dados simulado do modelo de Mistura Logístico com 30 ensaios Bernoulli

# Capítulo 4

## Aplicações

Para ilustrar a utilização dos modelos propostos, dois conjuntos de dados desafiadores na literatura de dados binários são considerados: Mortalidade dos besouros e Idade da primeira menstruação.

Os modelos bayesianos *Logístico* paramétrico (BP), paramétrico com transformação (BP TF), semi-paramétrico com mistura em escala (BSP ME), semi-paramétrico com transformação (BSP TF) e, por fim, semi-paramétrico com mistura em escala e posição (BSP MEP) são considerados.

### 4.1 Mortalidade dos Besouros

O banco de dados da mortalidade dos besouros consiste na contagem do número de besouros adultos mortos após a exposição ao gás dissulfeto de carbono durante cinco horas. O banco de dados está disponível em Bliss [1935] e segue abaixo,

**Tabela 4.1:** *Conjunto de dados: Mortalidade dos besouros*

Dose $CS_2$	Número de besouros expostos	Número de besouros mortos
49,09	59	6
52,99	60	13
56,91	62	18
60,84	56	28
64,76	63	52
68,69	59	53
72,61	62	61
76,54	60	60

Através do pacote *LaplacesDemon*, os modelos paramétricos e semi-paramétricos foram ajustados pelo algoritmo Metropolis-Hastings Adaptativo tal que uma cadeia de 10000 iterações foi utilizada na fase de adaptação com a estimação da matriz de covariâncias a cada 1000 iterações.

As condições para a obtenção da convergência e amostras não autocorrelacionadas estão apresentadas na Tabela 4.2 e foram avaliadas através de gráficos. Cabe ressaltar que a avaliação da eficiência computacional não foi um objetivo, deste modo, tentou-se estipular as mesmas condições para todos os modelos.

A distribuição a priori para  $\beta_0$  e  $\beta_1$  consideradas foram independentes e não-informativas ao defini-las como *Normal* com média 0 e precisão 0,01. Ainda assim, a informação a priori poderia ser elicitada para os coeficientes de regressão conforme a discussão do Capítulo 2.

Para os modelos que recorrem à utilização da transformação, a distribuição a priori de  $\lambda$  é definida como a distribuição *Normal*(1, 10), desta forma, é esperado uma curva de probabilidade simétrica a priori e acredita-se tanto na assimetria à esquerda como à direita. Note que a informação sobre  $\lambda$  poderia ser elicitada a partir da discussão da forma da curva de probabilidade.

Para o modelo de mistura em escala,  $H_0(\tau)$  corresponde a distribuição de *Kolmogorov-Smirnov* e para o modelo de mistura em locação e escala,  $H_0(\theta|\tau)$  é definida pela distribuição *Normal*(0,  $\tau$ ),

isto é, a curva de probabilidade esperada a priori é definida pela distribuição *Logística*. Para todos os modelos semi-paramétricos, foi estabelecido  $\alpha = 1$ , portanto, a crença de que a distribuição *Logística* seja apropriada aos dados é bastante fraca.

As estimativas dos modelos seguem na Tabela 4.3 e as medidas de distância seguem na Tabela 4.4. As estimativas a posteriori para  $\beta$  dos modelos semi-paramétricos não podem ser interpretadas através do conceito de dose letal mediana assim como no modelo paramétrico com transformação. Apesar desta limitação, é possível obter informações acerca da dose letal mediana uma vez que a curva de probabilidade está estimada.

Note que os modelos BP TF e BSP TF apresentam estimativas pontuais do parâmetro de assimetria  $\lambda$  maiores do que um que corresponde à presença de assimetria a direita assim como os intervalos de credibilidade 95%.

Os modelos simétricos BP e BSP ME são aqueles com os menores valores de SLOPC e os mais distantes dos dados, o que caracteriza um pior ajuste. Este é um resultado esperado uma vez que o conjunto de dados é conhecido pela assimetria à esquerda na literatura estatística.

Dentre os modelos assimétricos, o modelo BP TF é aquele que apresenta performance levemente superior, embora os pseudo-fatores de Bayes sejam bastante próximos a um: entre os modelos BP TF e BSP TF é igual a 1,007; BP TF e BSP MEP é igual a 1,046; BP TF e BSP ME é 1,087; BF e BP é 1,119. As reduções percentuais das medidas de distância e as as curvas de probabilidade apresentadas na Figura 4.1 conduzem a uma conclusão semelhante.

Em relação a comparação com o modelo *Complementar Log-log*, que é considerado o mais adequado a este conjunto de dados na literatura estatística, pode-se verificar que os modelos BP TF e BSP TF são superiores, segundo as medidas de distância. Por outro lado, a SLCPO favorece a regressão *Complementar Log-log*, apesar dos pseudo-fatores de Bayes estarem próximos a um: para o modelo BP TF é 1.030 e para o modelo BSP TF é 1.037.

**Tabela 4.2:** *Mortalidade dos besouros - Condições para o algoritmo MCMC*

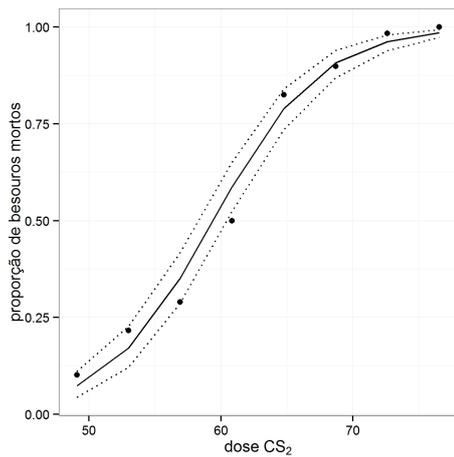
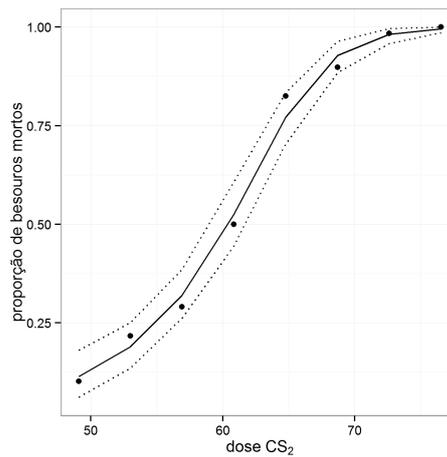
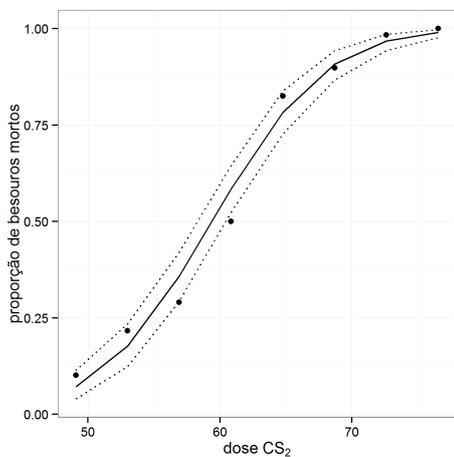
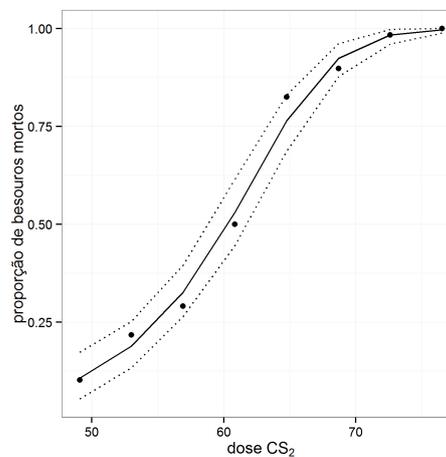
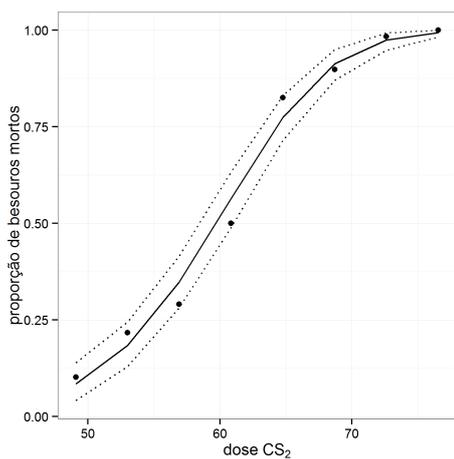
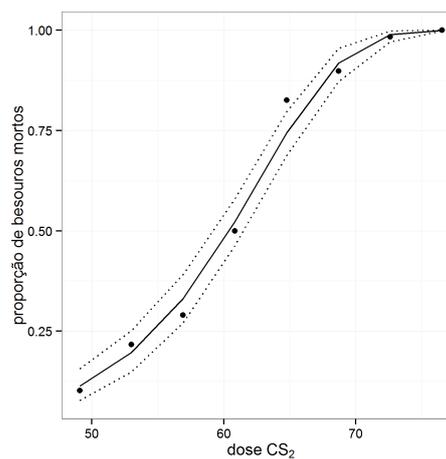
<b>Modelo</b>	<b>Aquecimento</b>	<b>Espaçamento</b>
BP	3000000	1000
BP TF	500000	1000
BSP ME	500000	1000
BSP TF	500000	1000
BSP MPS	500000	1000

**Tabela 4.3:** Mortalidade dos besouros - Estimativas para os modelos

Modelo	Estimativa	IC 95%	SLOPC
<i>Logístico</i> BP			-19,330
$\beta_0$	-14,797	(-17,199 ; -12,379)	
$\beta_1$	0,249	(0,210 ; 0,288)	
<i>Logístico</i> BP TF			-17,332
$\beta_0$	-14,886	(-17,634 ; -12,175)	
$\beta_1$	0,246	(0,203 ; 0,290)	
$\lambda$	1,376	(1,053 ; 1,738)	
<i>Logístico</i> BSP ME			-18,772
$\beta_0$	-9,900	(-14,718 ; -6,782)	
$\beta_1$	0,167	(0,114 ; 0,245)	
<i>Logístico</i> BSP TF			-17,623
$\beta_0$	-10,597	(-19,943 ; -6,619)	
$\beta_1$	0,176	(0,110 ; 0,324)	
$\lambda$	1,412	(0,998 ; 1,918)	
<i>Logístico</i> BSP MEP			-17,980
$\beta_0$	-10,689	(-15,528 ; -6,488)	
$\beta_1$	0,186	(0,127 ; 0,257)	
<i>Complementar Log-log</i> BP			-16,696
$\beta_0$	-9,764	(-11,293 ; -8,269)	
$\beta_1$	0,156	(0,132 ; 0,180)	

**Tabela 4.4:** Mortalidade dos besouros - Medidas de distância e redução percentual relativa

Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,00304 (0,0)	0,08769 (0,0)	0,01543 (0,0)
BP TF	0,00117 (61,5)	0,05400 (38,4)	0,00078 (94,9)
BSP ME	0,00306 (0,0)	0,08431 (3,8)	0,01528 (0,0)
BSP TF	0,0014 (47,5)	0,06055 (36,8)	0,00837 (43,3)
BSP MEP	0,00214 (30)	0,06393 (27,1)	0,01090 (29,3)
BP Complementar Log-log	0,00189 (38,2)	0,08098 (7,7)	0,01049 (32,0)

(a) *modelo BP*(b) *modelo BP TF*(c) *modelo BSP ME*(d) *modelo BSP TF*(e) *modelo BSP MEP*(f) *modelo Complementar Log-log BP*

**Figura 4.1:** Estimativas das curvas de probabilidade do conjunto de dados Mortalidade dos Besouros

## 4.2 Idade da primeira menstruação

O banco de dados da idade da primeira menstruação consiste na contagem do número de adolescentes que tiveram a primeira menstruação em uma determinada idade. O banco de dados está disponível em Milicer *et al.* [1968] e segue abaixo,

**Tabela 4.5:** *Conjunto de dados: Idade da primeira menstruação*

Idade	Número de adolescentes menstruadas	Número adolescentes entrevistadas
9,21	0	376
10,21	0	200
10,58	0	93
10,83	2	120
11,08	2	90
11,33	5	88
11,58	10	105
11,83	17	111
12,08	16	100
12,33	29	93
12,58	39	100
12,83	51	108
13,08	47	99
13,33	67	106
13,58	81	105
13,83	88	117
14,08	79	98
14,33	90	97
14,58	113	120
14,83	95	102
15,08	117	122
15,33	107	111
15,58	92	94
15,83	112	114
17,58	1049	1049

A mesma configuração da seção anterior foi considerada na utilização do algoritmo Metropolis-Hastings Adaptativo. As condições para convergência e amostras não autocorrelacionadas estão na Tabela 4.6. As mesmas distribuições a priori para os coeficientes de regressão e o parâmetro de assimetria  $\lambda$  da seção anterior também foram consideradas.

As escolhas de  $H_0$  e  $\alpha$  também são as mesmas da análise anterior. As estimativas dos modelos seguem na Tabela 4.7 e as medidas de distância seguem na Tabela 4.8.

**Tabela 4.6:** *Idade da primeira menstruação - Condições para o algoritmo MCMC em modelos Logístico*

Modelo	Aquecimento	Espaçamento
BP	1000000	20000
BP TF	500000	1000
BSP ME	500000	1000
BSP TF	3000000	1000
BSP MEP	1000000	10000
BP Stukel	2000000	10000

**Tabela 4.7:** Idade da primeira menstruação - Estimativas para os modelos bayesianos Logístico

Modelo	Estimativa	IC 95%	SLOPC
<i>Logístico</i> BP			-57.407
$\beta_0$	-21.157	(-22,722 ; -19,728)	
$\beta_1$	1.627	(1,518 ; 1,747)	
<i>Logístico</i> BP TF			-56.282
$\beta_0$	-21.481	(-23.062 ; -19.888)	
$\beta_1$	1.660	(1.537 ; 1.786)	
$\lambda$	0,881	(0,784 ; 0,984)	
<i>Logístico</i> BSP ME			-55,721
$\beta_0$	-13,172	(-17,812 ; -9,735)	
$\beta_1$	1,012	(0,748 ; 1,368)	
<i>Logístico</i> BSP TF			-52,341
$\beta_0$	-14,618	(-24,024 ; -10,356)	
$\beta_1$	1,131	(0,801 ; 1,863)	
$\lambda$	0,816	(0,665 ; 0,947)	
<i>Logístico</i> BSP MEP			-53,085
$\beta_0$	-15,398	(-20,139 ; -11,668)	
$\beta_1$	1,154	(0,863 ; 1,507)	
<i>Logístico</i> BP Stukel			-52.775
$\beta_0$	-15,398	(-20,139 ; -11,668)	
$\beta_1$	1,154	(0,863 ; 1,507)	
$\lambda_1$	0,122	(-0,017 ; 0,298)	
$\lambda_2$	0,386	(0,149 ; 0,714)	

**Tabela 4.8:** Idade da primeira menstruação - Medidas de distância e redução percentual relativa para os modelos Logístico

Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0,00164 (0,0)	0,06187 (0,0)	0,00928 (0,0)
BP TF	0,00166 (0,0)	0,08289 (-34,0)	0,00898 (5,4)
BSP ME	0,00147 (10,4)	0,0716 (-15,7)	0,00843 (9,2)
BSP TF	0,00136 (18,3)	0,08191 (-32,4)	0,00733 (21,4)
BSP MEP	0,00130 (20,7)	0,07057 (-14,1)	0,00716 (22,8)
BP Stukel	0,00123 (25,0)	0,07599 (-22,8)	0,00685 (26,2)

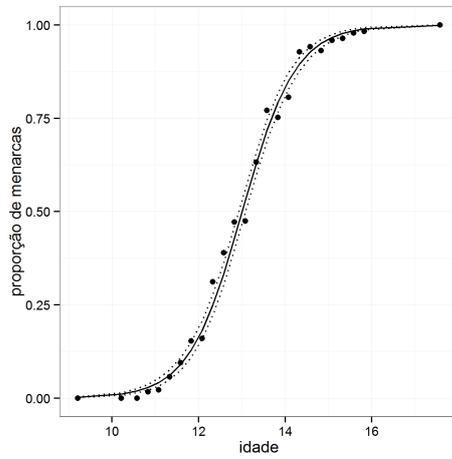
As estimativas a posteriori para  $\beta$  dos modelos semi-paramétricos não podem ser interpretadas através do conceito de razão de chances, assim como no modelo paramétrico com transformação. Apesar desta limitação, é possível obter informações acerca da razão de chance para duas idades, uma vez que a curva de probabilidade está estimada.

Note que os modelos BP TF e BSP TF apresentam estimativas pontuais do parâmetro de assimetria  $\lambda$  menores do que um, que representa existência de assimetria à esquerda. O modelo BP *Stukel* também indica assimetria à esquerda uma vez que o intervalo de credibilidade 95% não contém o valor zero para  $\lambda_2$ , porém, contém para  $\lambda_1$ .

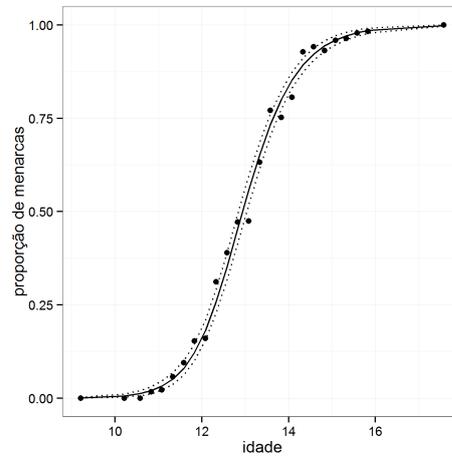
O modelo BP seguido pelos modelos BP TF e BSP ME são aqueles com os menores valores de SLOPC, o que é esperado já que tais modelos apresentam menos flexibilidade para lidar com discrepâncias da distribuição *Logística*. Enquanto, os modelos BSP TF e BSP MEP são aqueles com melhores performances: o pseudo-fator de Bayes entre BSP TF e BSP MEP é 1,023; entre BSP TF e BSP ME é 1,063; entre BSP TF e BP TF é 1,066; entre BSP TF e BP é 1,093.

A redução percentual das medidas de distância é crescente conforme o aumento da complexidade dos modelos, colocando o modelo BSP MEP como o mais indicado, entretanto, o modelo proposto

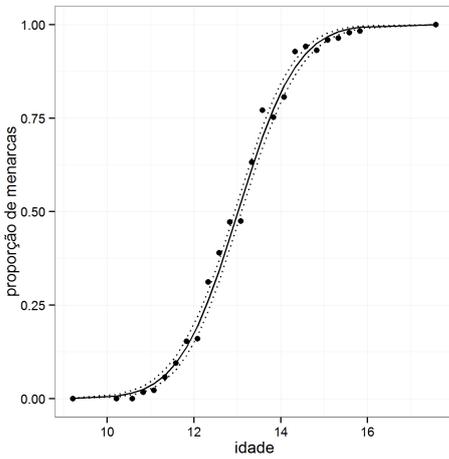
por [Stukel \[1988\]](#) ainda apresenta desempenho superior em todas as medidas de qualidade de ajuste, com exceção da medida de distância de *Kolmogorov-Smirnov*.



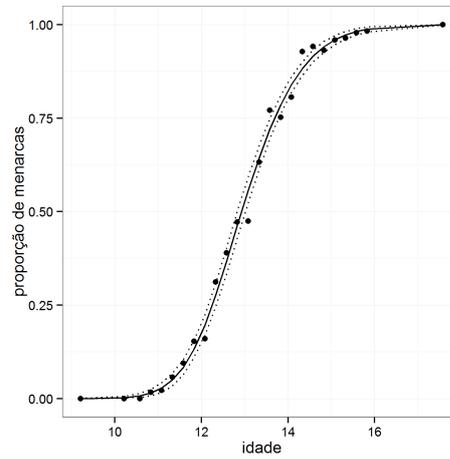
(a) modelo BP



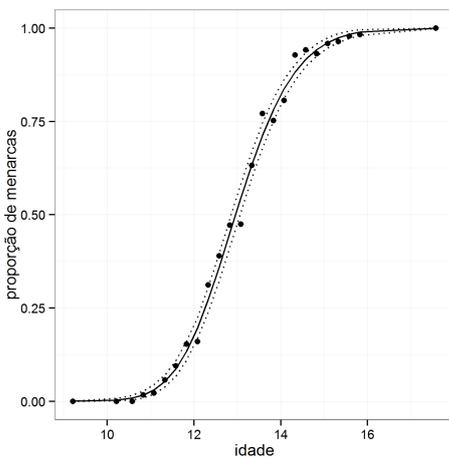
(b) modelo BP TF



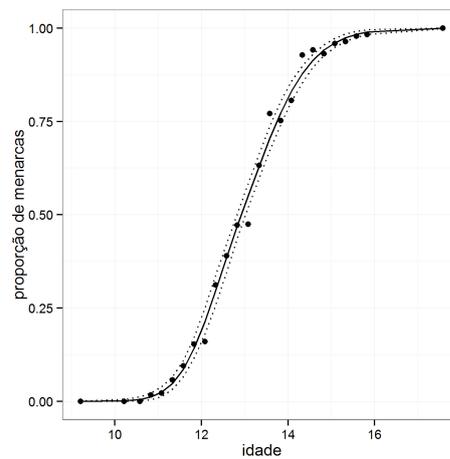
(c) modelo BSP ME



(d) modelo BSP TF



(e) modelo BSP MEP



(f) modelo BP Stukel

**Figura 4.2:** Estimativas das curvas de probabilidade do conjunto de dados Idade da primeira menstruação



## Capítulo 5

# Algoritmo: mistura em escala

No Capítulo 2, três modelos para lidar com discrepâncias do modelo *Logístico* foram apresentados sob a perspectiva bayesiana não-paramétrica. A abordagem bayesiana desde [Albert e Chib \[1993\]](#) tem usufruído de variáveis latentes como um artifício para a formulação de um algoritmo MCMC mais eficiente, porém, optou-se por não utilizá-las e o algoritmo de Metropolis-Hastings Adaptativo foi aplicado para a estimação dos modelos.

Desta forma, este capítulo tem por objetivo apresentar um algoritmo baseado em variáveis latentes e no Amostrador de Gibbs por Blocos introduzido por [Ishwaran e James \[2001\]](#) somente para o modelo de mistura em escala. Os resultados serão comparados com o ajuste deste mesmo modelo no capítulo anterior para o conjunto de dados Mortalidade dos Besouros que utiliza o algoritmo Metropolis Adaptativo.

Esta ilustração dará indícios da razão pela qual o algoritmo baseado em variáveis latentes parece não ser funcional no contexto bayesiano não-paramétrico.

### 5.1 Mistura em escala: variáveis latentes

Novamente, considere  $Y_{il}$  uma variável aleatória binária tal que  $Y_{il}|p_i \sim B(p_i)$  sendo  $p_i = F(\mathbf{x}'_i\boldsymbol{\beta}|0, \tau)$  para  $i = 1, \dots, L$  e  $l = 1, \dots, n_i$  e função de distribuição  $F$  é dita ser pertencente a uma família de posição e escala  $\mathbb{F}$ .

Como [Albert e Chib \[1993\]](#), define-se o conjunto de variáveis latentes  $W_{il} \sim N(\mathbf{x}'_i\boldsymbol{\beta}, \tau)$ . Portanto, pode-se reescrever as variáveis binárias como a discretização de tais variáveis latentes,

$$Y_{il} = I_{\{W_{il} > 0\}},$$

pois,

$$\begin{aligned} p_i &= P(Y_{il} = 1) \\ &= P(W_{il} > 0 | \mu = \mathbf{x}'_i\boldsymbol{\beta}, \tau) \\ &= P(W_{il} - \mathbf{x}'_i\boldsymbol{\beta} > -\mathbf{x}'_i\boldsymbol{\beta} | \mu = 0, \tau) \\ &= 1 - P(W_{il} - \mathbf{x}'_i\boldsymbol{\beta} \leq -\mathbf{x}'_i\boldsymbol{\beta} | \mu = 0, \tau) \\ &= P(W_{il} - \mathbf{x}'_i\boldsymbol{\beta} \leq \mathbf{x}'_i\boldsymbol{\beta} | \mu = 0, \tau) \\ &= F(\mathbf{x}'_i\boldsymbol{\beta}|0, \tau) \end{aligned} \tag{5.1}$$

No modelo de mistura em escala, a função de densidade derivada da função de distribuição  $F$  é dada por

$$f(W|Y, \mathbf{x}'\boldsymbol{\beta}, H) = \int \phi(W|\mathbf{x}'\boldsymbol{\beta}, \tau) dH(\tau). \tag{5.2}$$

Portanto, a função de verossimilhança é definida por

$$L(\boldsymbol{\beta}, H|\mathbf{Y}, \mathbf{X}) = \prod_{i=1}^n \prod_{l=1}^{n_i} p_i^{y_{il}} (1 - p_i)^{(1-y_{il})}$$

sendo que  $p_i = F(\mathbf{x}'\boldsymbol{\beta}|0, \tau)$  para  $\tau \sim H$ .

Considerando as distribuições a priori discutidas no Capítulo 2,

$$\begin{aligned} \boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \mathbf{T}_\beta &\sim N_p(\boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ H|\alpha &\sim P(\alpha, H_0), \end{aligned} \quad (5.3)$$

sendo  $H_0$  a distribuição base do processo de *Dirichlet* que é definida a partir da escolha da distribuição  $F_0$ .

O passo seguinte é o cálculo da distribuição a posteriori

$$\pi(\boldsymbol{\beta}, H|\mathbf{Y}, \mathbf{X}) \propto L(\boldsymbol{\beta}, H|\mathbf{Y}, \mathbf{X})\pi(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \mathbf{T}_\beta)P(\alpha, H_0|v),$$

que pode ser construída a partir da amostragem das distribuições condicionais completas a posteriori,

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, H, \boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ \pi(H|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}). \end{aligned}$$

Como já dito, a amostragem da distribuição a posteriori de  $H$  não é simples, contudo, há algoritmos que podem ser utilizados para dados binários através das variáveis latentes.

## 5.2 Amostradores de Gibbs

Há duas abordagens computacionais para lidar com misturas de processos de Dirichlet que são denominados como Amostrador de Gibbs da Urna de Pólia e Amostrador de Gibbs em Blocos por [Ishwaran e James \[2001\]](#).

A primeira abordagem foi apresentada por [Escobar \[1994\]](#), detalhada em [Escobar e West \[1995\]](#) e, mais tarde, refinamentos foram apresentados em [MacEachern \[1994\]](#), [Escobar e West \[1998b\]](#), [Escobar e West \[1998a\]](#) e [MacEachern e Müller \[1998\]](#).

A segunda abordagem foi introduzida por [Ishwaran e Zarepour \[2000\]](#) como um algoritmo amostrador de Gibbs simples e geral que permite a amostragem direta de valores provenientes da distribuição a posteriori da medida aleatória  $H$ . Mais tarde, [Ishwaran e James \[2001\]](#) apresentam mais detalhes e formalizaram o procedimento.

Esta última abordagem será considerada neste texto pela facilidade em trabalhar nas situações em que não há conjugação entre a distribuição da variável latente  $W$ , definida pelas componentes da mistura *Normal*, e a distribuição de mistura  $H_0$ .

Para aplicar tal abordagem, é necessário considerar um truncamento do processo Dirichlet associado à medida aleatória  $H$  e, desta forma, pode-se reescrever o modelo (5.2) - (5.3) ao considerar  $\tau(u_i) = \tau(Z_{K_i})$  tal que  $K_i$  para  $i = 1, \dots, L$  são variáveis de classificação para identificar a variável  $u_i$ , que define  $\tau$ , associada a cada possível valor de  $Z$ . Então,

$$\begin{aligned} W_{il}|\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}_\tau, \mathbf{K} &\stackrel{i.i.d.}{\sim} \phi(W_{il}|\mathbf{x}'_i\boldsymbol{\beta}, \tau(Z_{K_i\tau})) \quad \text{para } l = 1, \dots, n_i \text{ e } i = 1, \dots, L, \\ \boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \tau_\beta &\sim N_p(\boldsymbol{\mu}_\beta, \mathbf{T}_\beta), \\ Z_{j\tau}|v &\stackrel{i.i.d.}{\sim} H_0(v) \quad \text{para } j = 1, \dots, N, \\ \mathbf{K}|\mathbf{q} &\sim \text{Multinomial}(L, q_1, \dots, q_N), \\ \mathbf{q}|\alpha &\sim \text{Dirichlet Generalizada}_N(\alpha, \mathbf{1}), \end{aligned} \quad (5.4)$$

Este truncamento do processo de Dirichlet é diferente de considerar um processo de Dirichlet finito como no Capítulo 2, pois a escolha do número de componentes  $N$  é feita de forma que as componentes desprezadas sejam estimadas iguais a zero.

A questão do número de componentes pode ser respondida a partir do princípio de que as densidades marginais  $f(W)$  e  $f_N(W)$  devem ser similares com  $f_N(W)$  denotando a densidade de  $W$  quando o processo de Dirichlet truncado é considerado. Neste sentido, Ishwaran e James [2001] apresentam o limite,

$$\|f(\mathbf{W}) - f_N(\mathbf{W})\|_1 \leq 4n \exp(-(N-1)/\alpha), \quad (5.5)$$

Ao reescrever o modelo na forma (5.4), também é possível reescrever a distribuição a posteriori de  $H|\mathbf{W}, \boldsymbol{\beta}, \mathbf{X}, v$ ,

$$\pi(H|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, v) = \pi(\mathbf{Z}_\tau, \mathbf{K}, \mathbf{q}, \alpha|\boldsymbol{\beta}, \mathbf{W}, \mathbf{X}, v).$$

E pode-se construir tal distribuição a posteriori através da amostragem das distribuições condicionais completas,

$$\begin{aligned} & \pi(\mathbf{Z}_\tau|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v), \\ & \pi(\mathbf{K}|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}_\tau, \mathbf{q}), \\ & \pi(\mathbf{q}|\mathbf{K}, \alpha). \end{aligned}$$

### 5.2.1 Amostragem de $\mathbf{Z}_\tau|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v$

A obtenção da distribuição condicional completa a posteriori  $\mathbf{Z}_\tau|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v$  ocorre em duas etapas. Para tanto, define-se o conjunto  $K^* = \{K_1^*, \dots, K_m^*\}$  que consiste nos valores únicos do vetor  $\mathbf{K}$ . Então,

$$\begin{aligned} \pi(\mathbf{Z}_\tau|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v) & \propto \prod_{\{k \in \{1, \dots, N\} - K^*\}} h_0(Z_{k\tau}|v) \times \\ & \times \prod_{j=1}^m \left( h_0(Z_{k_j^* \tau}|v) \prod_{\{i: K_i = k_j^*\}} \prod_{l=1}^{n_i} \phi(W_{il}|\mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k_j^* \tau})) \right), \end{aligned} \quad (5.6)$$

isto é, a amostragem de  $\mathbf{Z}_\tau|\mathbf{K}, \boldsymbol{\beta}, \mathbf{W}, v$  corresponde às etapas,

- i.  $Z_{k\tau} \stackrel{i.i.d.}{\sim} h_0(k\tau|v)$  para  $k \in \{1, \dots, N\} - K^*$ ;
- ii.  $Z_{k\tau} \stackrel{ind}{\sim} h_0(k\tau|v) \prod_{\{i: K_i = k\}} \prod_{l=1}^{n_i} \phi(W_{il}|\mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k\tau}))$  para  $k \in K^*$ .

As mesmas escolhas para  $H_0$  discutidas no Capítulo 2 podem ser implementadas, pois a inserção do algoritmo Metropolis-Hastings no Amostrador de Gibbs quando  $H_0$  é a distribuição de *Kolmogorov-Smirnov* não é difícil.

### *Cauchy, t-student e Normal*

Para  $\tau(Z_\tau) = Z_\tau$  sendo  $Z_\tau \sim \text{Gama}(v/2, v/2)$  com densidade,

$$h_0(Z_\tau) = \frac{(v/2)^{\frac{v}{2}}}{\Gamma(v/2)} Z_\tau^{\frac{v}{2}-1} \exp\left\{-\frac{v}{2} Z_\tau\right\} \mathbb{1}_{(Z_\tau > 0)},$$

segue que  $F_0$  é dada pela distribuição  $t_v(0, 1)$ . Se  $v = 1$ ,  $F_0$  é a distribuição *Cauchy*(0, 1), enquanto se  $v \rightarrow \infty$ ,  $F_0$  é a distribuição *Normal*(0, 1) sendo que para  $v > 30$ , a distribuição *t-student* com  $v$  graus de liberdade já é considerada uma boa aproximação para a distribuição *Normal*.

Para a implementação do algoritmo apresentado na seção anterior, é necessário estabelecer a amostragem da distribuição a posteriori de  $Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v$  para  $k \in K^*$  referente ao passo *ii*, isto é,

$$\begin{aligned}
\pi(Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v) &\propto \\
&\propto h_0(Z_{k\tau} | v) \prod_{\{i: K_i=k\}} \prod_{\{l=1\}}^{n_i} \phi(W_{il} | \mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k\tau})) \\
&\propto Z_{k\tau}^{v/2-1} \exp\left\{-Z_{k\tau} \frac{v}{2}\right\} \prod_{\{i: K_i=k\}} \prod_{\{l=1\}}^{n_i} Z_{k\tau}^{1/2} \exp\left\{-\frac{Z_{k\tau}}{2} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\} \\
&\propto Z_{k\tau}^{v/2-1} \exp\left\{-Z_{k\tau} \frac{v}{2}\right\} Z_{k\tau}^{m_k/2} \exp\left\{-\frac{Z_{k\tau}}{2} \sum_{\{i: K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\} \\
&\propto Z_{k\tau}^{v/2+m_k/2-1} \exp\left\{-\frac{Z_{k\tau}}{2} \left[ \sum_{\{i: K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 + v \right]\right\}, \tag{5.7}
\end{aligned}$$

tal que  $m_k = \sum_{\{i: K_i=k\}} n_i$ , ou seja, a cardinalidade do conjunto  $\{i : K_i = k\}$  para  $k \in K^*$ .

É possível identificar a equação (5.7) como o núcleo da distribuição  $Gama(v_1^W, v_2^W)$  com parâmetros dados por

$$\begin{aligned}
v_1^W &= v/2 + m_k/2, \\
v_2^W &= \frac{1}{2} \sum_{\{i: K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 + \frac{v}{2}.
\end{aligned}$$

### Logística

Para obter a distribuição  $Logística(0, 1, \lambda)$ , deve-se considerar

$$\tau(Z_\tau) = \left(\frac{1}{2Z_\tau}\right)^2 \quad \text{tal que} \quad Z_\tau \sim \text{Kolmogorov-Smirnov},$$

em que a distribuição de  $Kolmogorov-Smirnov$ .

A amostragem da distribuição a posteriori de  $Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}$  para  $k \in K^*$  é dada por

$$\begin{aligned}
\pi(Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}) &\propto \\
&\propto h_0(Z_{k\tau}) \prod_{\{i: K_i=k\}} \prod_{l=1}^{n_i} \phi(W_{il} | \mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k\tau})) \\
&\propto 8 \sum_{k=1}^{\infty} (-1)^{n+1} n^2 Z_{k\tau} \exp\{-2n^2 Z_{k\tau}^2\} \prod_{\{i: K_i=k\}} \prod_{l=1}^{n_i} \left(\frac{1}{4Z_{k\tau}^2}\right)^{1/2} \exp\left\{-\frac{1}{8Z_{k\tau}^2} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\} \\
&\propto 8 \sum_{k=1}^{\infty} (-1)^{n+1} n^2 Z_{k\tau} \exp\{-2n^2 Z_{k\tau}^2\} \left(\frac{1}{4Z_{k\tau}^2}\right)^{m_k/2} \\
&\quad \times \exp\left\{-\frac{1}{8Z_{k\tau}^2} \sum_{\{i: K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\}, \tag{5.8}
\end{aligned}$$

que resulta em uma distribuição a posteriori não conhecida, contudo, para tal distribuição, [Chen e Dey \[1998\]](#) sugerem uma distribuição proposta para a implementação do algoritmo de Metropolis-Hastings.

Chen e Dey [1998] constroem uma distribuição proposta através dos resultados empíricos de Albert e Chib [1993], nos quais uma variável com distribuição  $t_v$  é aproximadamente  $b$  vezes uma variável com distribuição *Logística*, para escolhas apropriadas de  $v$  e  $b$ .

Este é o ponto de partida para Chen e Dey [1998] que definem uma aproximação similar através da mistura em escala de distribuições *Normal* tal que a igualdade nos dois primeiros momentos entre as distribuições  $t_v$  e *Logística* seja satisfeita com a restrição de que as caudas da distribuição proposta sejam mais pesadas do que a distribuição de *Kolmogorov-Smirnov*, desta forma, é encontrado que as melhores escolhas são  $v = 5$  e  $b = 0.712$ .

Como consequência de tal aproximação, a distribuição de *Kolmogorov-Smirnov* é aproximada por,

$$h_0^a(Z_\tau) = \frac{(v/8b^2)^{(v/2)}}{\Gamma(v/2)} (Z_\tau^2)^{-(v/2+1)} \exp\left\{-\frac{v}{8b^2 Z_\tau^2}\right\} 2Z_\tau, \quad (5.9)$$

que é equivalente à atribuição da distribuição *Inversa Gama*( $v/2, v/8b^2$ ) para  $Z^2$ .

Então, pode-se construir uma densidade proposta para a distribuição dada na equação (5.8) que consiste em uma distribuição a posteriori calculada, considerando a distribuição base  $h_0^a(Z)$  dada na equação (5.9),

$$\begin{aligned} \pi^a(Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}) &= \\ &= h_0^a(Z_{k\tau}) \prod_{\{i:K_i=k\}} \prod_{l=1}^{n_i} \phi(W_{il} | \mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k\tau})) \\ &= \frac{(v/8b^2)^{(v/2)}}{\Gamma(v/2)} (Z_{k\tau}^2)^{-(v/2+1)} \exp\left\{-\frac{v}{8b^2 Z_{k\tau}^2}\right\} 2Z_{k\tau} \\ &\quad \times \prod_{\{i:K_i=k\}} \prod_{l=1}^{n_i} \left(\frac{1}{4Z_{k\tau}^2}\right)^{1/2} \exp\left\{-\frac{1}{8Z_{k\tau}^2} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\} \\ &\propto (Z_{k\tau}^2)^{-(v/2+1)} \exp\left\{-v/8b^2 Z_{k\tau}^2\right\} Z_{k\tau} \\ &\quad \times \left(\frac{1}{4Z_{k\tau}^2}\right)^{m_k/2} \exp\left\{-\frac{1}{8Z_{k\tau}^2} \sum_{\{i:K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right\} \\ &\propto (Z_{k\tau}^2)^{-(v/2+m_k/2+1)} \\ &\quad \times \exp\left\{-\frac{1}{8Z_{k\tau}^2} \left[\frac{v}{b^2} + \sum_{\{i:K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2\right]\right\} Z_{k\tau}, \end{aligned}$$

sendo que tal distribuição é uma derivação da distribuição para  $Z_{k\tau}^2 | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}$  dada pela *Inversa Gama*( $v_1^W, v_2^W$ ) com

$$\begin{aligned} v_1^W &= v/2 + m_k/2, \\ v_2^W &= \frac{1}{8} \left[ \frac{v}{b^2} + \sum_{\{i:K_i=k\}} \sum_{\{l=1\}}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 \right]. \end{aligned}$$

E a probabilidade de aceitação de  $Z_{k\tau}^*$  gerado pela distribuição proposta é dada por

$$\begin{aligned} \rho &= \frac{\pi(Z_{k\tau}^* | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v) / \pi^a(Z_{k\tau}^* | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v)}{\pi(Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v) / \pi^a(Z_{k\tau} | \mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{K}, v)} \\ &= \frac{h_0(Z_{k\tau}^*) / h_0^a(Z_{k\tau}^*)}{h_0(Z_{k\tau}) / h_0^a(Z_{k\tau})}. \end{aligned}$$

Note que esta probabilidade de aceitação exige a avaliação da densidade da distribuição de

*Kolmogorov-Smirnov* e, como já dito, somente é possível avaliá-la aproximadamente. A escolha do índice de truncamento do somatório segue a discussão do Capítulo 2 e é dada pela equação (2.10).

### 5.2.2 Amostragem de $\mathbf{K}|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{q}$

Por sua vez, note que a distribuição  $\mathbf{K}|\mathbf{q}$  é definida pela distribuição *Multinomial*( $q_1, \dots, q_N$ ) que pode ser reescrita para cada componente do vetor  $\mathbf{K}$  na forma

$$K_i \sim \sum_{k=1}^N \mathbb{1}_{(K_i=k)} q_k \quad \text{para } i = 1, \dots, L. \quad (5.10)$$

É necessário escrever a distribuição de  $\mathbf{W}|\boldsymbol{\beta}, \tau$  como uma distribuição multivariada, isto é,  $N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{T}_Z)$  tal que  $\mathbf{T}$  é uma matriz diagonal com o  $i$ -ésimo elemento dado por  $\tau(Z_{K_i})$ .

A partir da equação (5.10) tem-se que

$$\begin{aligned} \pi(\mathbf{K}|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{q}) &\propto \pi(\mathbf{K}|\mathbf{q}) \phi_n(\mathbf{W}|\mathbf{X}\boldsymbol{\beta}, \mathbf{T}_Z) \\ &\propto \prod_{i=1}^L \left( \sum_{k=1}^N \mathbb{1}_{(K_i=k)} q_k \right) \prod_{i=1}^L \prod_{l=1}^{n_i} \phi(W_{il}|\mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{K_i})) \\ &\propto \prod_{i=1}^L \left( \sum_{k=1}^N \mathbb{1}_{(K_i=k)} q_k \prod_{l=1}^{n_i} \phi(W_{il}|\mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{K_i})) \right) \\ &\propto \prod_{i=1}^L \left( \sum_{k=1}^N \mathbb{1}_{(K_i=k)} q_{i,k} \right), \end{aligned} \quad (5.11)$$

tal que

$$\begin{aligned} q_{i,k} &\propto q_k \prod_{l=1}^{n_i} \phi(W_{il}|\mathbf{x}'_i \boldsymbol{\beta}, \tau(Z_{k\tau})) \\ &\propto q_k \prod_{l=1}^{n_i} \tau(Z_{k\tau})^{1/2} \exp \left\{ -\frac{\tau(Z_{k\tau})}{2} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 \right\} \\ &\propto q_k \tau(Z_{k\tau})^{n_i/2} \exp \left\{ -\frac{\tau(Z_{k\tau})}{2} \sum_{l=1}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 \right\}, \end{aligned} \quad (5.12)$$

isto é,  $K_i|\mathbf{W}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{q} \sim \text{Multinomial}(1, q_{i,1}, \dots, q_{i,N})$  para  $i = 1, \dots, L$ .

Então,  $q_{i,k}$  para  $i = 1, \dots, L$  é definido para escolha possível de  $F_0$ ,

#### *Cauchy, t-student e Normal*

$$q_{i,k} \propto q_k \tau(Z_{k\tau})^{n_i/2} \exp \left\{ -\frac{\tau(Z_{k\tau})}{2} \sum_{l=1}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 \right\};$$

#### *Logística*

$$q_{i,k} \propto q_k \left( \frac{1}{4Z_{k\tau}^2} \right)^{n_i/2} \exp \left\{ -\frac{1}{8Z_{k\tau}^2} \sum_{l=1}^{n_i} (W_{il} - \mathbf{x}'_i \boldsymbol{\beta})^2 \right\};$$

### 5.2.3 Amostragem de $\mathbf{q}|\mathbf{K}, \alpha$

Wong [1998] apresenta a distribuição *Dirichlet Generalizada*( $\mathbf{a}, \mathbf{b}$ ) como a distribuição conjugada para a distribuição *Multinomial* atribuída a  $\mathbf{K}$ . Deste modo, a distribuição condicional completa a posteriori  $\mathbf{q}|\alpha, \mathbf{K}$  é também dada pela distribuição *Dirichlet Generalizada*( $\mathbf{a}^{\mathbf{W}}, \mathbf{b}^{\mathbf{W}}$ ) tal que

$$\begin{aligned} a_k^{\mathbf{W}} &= 1 + m_k, \\ b_k^{\mathbf{W}} &= \alpha + \sum_{j=k+1}^N m_j. \end{aligned} \quad (5.13)$$

relembrando que  $m_k$  o número de  $K_i$  iguais a  $k$  para  $k = 1, \dots, N-1$  e  $i = 1, \dots, L$ .

Cabe ressaltar que para gerar  $\mathbf{q}|\alpha, \mathbf{K}$  é preferível gerar  $V_k \sim \text{Beta}(a_k^{\mathbf{W}}, b_k^{\mathbf{W}})$  para  $k = 1, \dots, N-1$  considerando  $V_N = 1$  e definir

$$\begin{aligned} q_1 &= V_1, \\ q_k &= (1 - V_1)(1 - V_2) \times \dots \times (1 - V_{k-1})V_k \quad \text{para } k = 1, \dots, N-1. \end{aligned}$$

### 5.2.4 Amostragem de $\beta|\mathbf{W}, \mathbf{X}, \mathbf{Z}, \mathbf{K}, \mu_\beta, \mathbf{T}_\beta$

Para a amostragem da distribuição condicional completa a posteriori  $\pi(\beta|\mathbf{W}, \mathbf{X}, \mathbf{Z}, \mathbf{K}, \mu_\beta, \mathbf{T}_\beta)$  deve-se considerar a densidade de  $\mathbf{W}$  na forma matricial. Deste modo,

$$\begin{aligned} \pi(\beta|\mathbf{W}, \mathbf{X}, \mathbf{Z}, \mathbf{K}, \mu_\beta, \tau_\beta) &= \pi(\beta|\tau_\beta)\phi_n(\mathbf{W}|\mathbf{X}\beta, \mathbf{T}_\mathbf{Z}) \\ &= (2\pi)^{\binom{-p}{2}}|\mathbf{T}_\beta| \exp\left\{-\frac{1}{2}(\beta - \mu_\beta)'\mathbf{T}_\beta(\beta - \mu_\beta)\right\} \\ &\quad \times (2\pi)^{\binom{-n}{2}}|\mathbf{T}_\mathbf{Z}| \exp\left\{-\frac{1}{2}(\mathbf{W} - \mathbf{X}\beta)'\mathbf{T}_\mathbf{Z}(\mathbf{W} - \mathbf{X}\beta)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\beta - \mu_\beta)'\mathbf{T}_\beta(\beta - \mu_\beta)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{W} - \mathbf{X}\beta)'\mathbf{T}_\mathbf{Z}(\mathbf{W} - \mathbf{X}\beta)\right\} \end{aligned} \quad (5.14)$$

Olhando somente o expoente na equação (5.14),

$$\begin{aligned} &(\beta - \mu_\beta)'\mathbf{T}_\beta(\beta - \mu_\beta) + (\mathbf{W} - \mathbf{X}\beta)'\mathbf{T}_\mathbf{Z}(\mathbf{W} - \mathbf{X}\beta) = \\ &= (\beta' - \mu_\beta')\mathbf{T}_\beta(\beta - \mu_\beta) + (\mathbf{W}' - \beta'\mathbf{X}')\mathbf{T}_\mathbf{Z}(\mathbf{W} - \mathbf{X}\beta) \\ &= \beta'\mathbf{T}_\beta\beta - \beta'\mathbf{T}_\beta\mu_\beta - \mu_\beta'\mathbf{T}_\beta\beta + \mu_\beta'\mathbf{T}_\beta\mu_\beta + \beta'\mathbf{T}_\beta\beta \\ &\quad + \mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{W} - \mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{W} + \beta'\mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X}\beta \\ &\propto \beta'\mathbf{T}_\beta\beta - 2\mu_\beta'\mathbf{T}_\beta\beta - 2\mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X}\beta \\ &\propto \beta'(\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})\beta - 2(\mu_\beta'\mathbf{T}_\beta + \mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{X})\beta \end{aligned} \quad (5.15)$$

$$\begin{aligned} &\propto \beta'(\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})\beta - 2(\mu_\beta'\mathbf{T}_\beta + \mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{X})\beta \\ &\quad + (\mu_\beta'\mathbf{T}_\beta + \mathbf{W}'\mathbf{T}_\mathbf{Z}\mathbf{X})(\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})^{-1}(\mu_\beta\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{W}) \end{aligned} \quad (5.16)$$

$$\begin{aligned} &\propto \left(\beta - (\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})^{-1}(\mu_\beta\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{W})\right)'(\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X}) \\ &\quad \times \left(\beta - (\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})^{-1}(\mu_\beta\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{W})\right). \end{aligned} \quad (5.17)$$

isto é,  $\beta|\mathbf{W}, H, \tau_\beta \sim N_p(\mu_\beta^{\mathbf{W}}, \mathbf{T}_\beta^{\mathbf{W}})$  com parâmetros,

$$\mu_\beta^{\mathbf{W}} = (\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{X})^{-1}(\mu_\beta\mathbf{T}_\beta + \mathbf{X}'\mathbf{T}_\mathbf{Z}\mathbf{W}),$$

$$\mathbf{T}_Z \beta^W = (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}). \quad (5.18)$$

A equação (5.16) pode ser justificada pela necessidade de completar a forma quadrática na equação (5.15) a fim de obter um núcleo de uma distribuição *Normal* multivariada, dado que a distribuição a priori  $\pi(\beta | \mu_\beta, \tau_\beta)$  é conjugada da distribuição da variável latente  $W$ . Para tanto, considere uma matriz  $\mathbf{A}$  desconhecida tal que

$$\begin{aligned} (\beta - \mathbf{A})' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) (\beta - \mathbf{A}) &= \\ &= (\beta' - \mathbf{A}') (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) (\beta - \mathbf{A}) \\ &= \beta' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \beta - \beta' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \mathbf{A} \\ &\quad - \mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \beta + \mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \mathbf{A} \\ &= \beta' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \beta - 2\mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \beta \\ &\quad + \mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \mathbf{A}. \end{aligned} \quad (5.19)$$

Comparando as equações (5.15) e (5.19) segue que

$$\begin{aligned} (\mu_\beta' \mathbf{T}_\beta + \mathbf{W}' \mathbf{T}_Z \mathbf{X}) \beta &= \mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \beta \\ (\mu_\beta' \mathbf{T}_\beta + \mathbf{W}' \mathbf{T}_Z \mathbf{X}) &= \mathbf{A}' (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X}) \\ \mathbf{A}' &= (\mu_\beta' \mathbf{T}_\beta + \mathbf{W}' \mathbf{T}_Z \mathbf{X}) (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X})^{-1} \\ \mathbf{A} &= \left( (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X})^{-1} \right)' (\mu_\beta \mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{W}) \\ &= (\mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{X})^{-1} (\mu_\beta \mathbf{T}_\beta + \mathbf{X}' \mathbf{T}_Z \mathbf{W}). \end{aligned} \quad (5.20)$$

Finalmente, basta substituir a equação (5.20) em (5.19) para obter a equação (5.17).

### 5.2.5 Amostragem de $\mathbf{W} | \mathbf{Y}, \mathbf{X}, \beta, \mathbf{Z}_\tau, \mathbf{K}$

Para a amostragem de  $\pi(\mathbf{W} | \beta, \mathbf{Z}_\tau, \mathbf{K}, \mathbf{Y})$  que corresponde à amostragem de  $\phi(\mathbf{W} | \beta, \mathbf{Z}_\tau, \mathbf{K}, \mathbf{Y})$  truncada por  $\mathbf{Y}$ . Segue-se Robert [1995],

$$\pi(\mathbf{W} | \mathbf{Y}, \mathbf{X}, \beta, \mathbf{Z}_\tau, \mathbf{K}) \propto \prod_{i=1}^L \prod_{l=1}^{n_i} \phi(W_{il} | \mathbf{x}_i' \beta, \tau(Z_{K_i \tau})) \mathbb{1}_{(Y_{il} = \mathbb{1}_{(W_{il} > 0)})}. \quad (5.21)$$

### 5.2.6 Algoritmo

Portanto, a amostragem de  $\pi(\mathbf{W}, \beta, H | \mathbf{Y}, \mathbf{X})$  pode ser realizada a partir do seguinte algoritmo,

1. Defina valores iniciais para  $\mathbf{W}$ ,  $\mathbf{Z}_\tau$ ,  $\mathbf{K}$ ,  $\mathbf{q}$ ,  $\alpha$  e  $\beta$ ;
2. Gere  $H$  proveniente de  $\pi(H | \mathbf{W}, \mathbf{X}, \beta)$  a partir das partes,
  - i.  $\mathbf{Z}_\tau \sim \pi(\mathbf{Z}_\tau | \mathbf{W}, \mathbf{X}, \beta, \mathbf{K}, v)$  definida na equação (5.6);
  - ii.  $\mathbf{K} \sim \pi(\mathbf{K} | \mathbf{W}, \mathbf{X}, \beta, \mathbf{Z}_\tau, \mathbf{q})$  detalhada nas equações (5.11) - (5.12);
  - iii.  $\mathbf{q} \sim \text{Dirichlet Generalizada}_N(\mathbf{a}^{\mathbf{W}}, \mathbf{b}^{\mathbf{W}})$  com parâmetros definidos na equação (5.13);
3.  $\beta \sim N_p(\mu_\beta^W, \mathbf{T}_\beta^W)$  com parâmetros apresentados na equação (5.18);
4. Gere  $\mathbf{W} \sim \pi(\mathbf{W} | \mathbf{Y}, \mathbf{X}, \beta, \mathbf{Z}_\tau, \mathbf{K})$  expressa na equação (5.21).

Observe que no primeiro passo do algoritmo, há a necessidade de introduzir valores iniciais. Tais valores podem ser facilmente escolhidos a partir das distribuições a priori apresentadas em (5.4).

### 5.3 Ilustração comparativa dos algoritmos

É esperado que a estimação resultante do algoritmo baseado em variáveis latentes seja similar ao algoritmo Metropolis-Hastings Adaptativo. Desta forma, o modelo semi-paramétrico de mistura em escala foi ajustado por essas duas vias, considerando o conjunto de dados da Mortalidade dos Besouros e os resultados foram comparados. O objetivo é apresentar subsídios para a discussão de que, possivelmente, o algoritmo de variáveis latentes como introduzido por Albert e Chib [1993] não seja adequado na abordagem bayesiana não-paramétrica.

Para tanto, a distribuição a priori para  $\beta_0$  e  $\beta_1$  consideradas foram independentes e não-informativas ao defini-las como *Normal* com média 0 e precisão 0,01. Para o processo *Dirichlet*,  $H_0(\tau)$  consiste na distribuição de Kolmogorov-Smirnov, isto é, a curva de probabilidade esperada a priori é definida pela distribuição *Logística* e o parâmetro de  $\alpha = 1$ , portanto, a crença de que a distribuição *Logística* seja apropriada aos dados é bastante fraca.

Os resultados dos modelos bayesianos *Logístico* paramétrico (BP) e semi-paramétrico com mistura em escala (BSP ME) do capítulo anterior são considerados novamente, junto ao modelo bayesiano semi-paramétrico com mistura em escala considerando variáveis latentes (BSP MEVL) discutido neste capítulo.

As condições para a implementação do MCMC estão na Tabela 5.1 e pode-se observar claramente que o algoritmo que usufrui de variáveis latentes é muito mais eficiente do que o Metropolis-Hastings Adaptativo para este banco de dados. As estimativas a posteriori estão na Tabela 5.2.

**Tabela 5.1:** Ilustração comparativa dos algoritmos - Condições para o algoritmo MCMC em modelos *Logístico*

Modelo	Aquecimento	Espaçamento
BP	3000000	1000
BSP ME	500000	1000
BSP ME	25000	250

**Tabela 5.2:** Ilustração comparativa dos algoritmos - Estimativas para os modelos bayesianos *Logístico*

Modelo	Estimativa	IC 95%
<i>Logístico</i> BP		
$\beta_0$	-14.797	(-17.199 ; -12.379)
$\beta_1$	0.249	(0.210 ; 0.288)
<i>Logístico</i> BSP ME		
$\beta_0$	-9.900	(-14.718 ; -6.782)
$\beta_1$	0.167	(0.114 ; 0.245)
<i>Logístico</i> BSP MEVL		
$\beta_0$	-16.957	(-28.969 ; -9.539)
$\beta_1$	0.286	(0.162 ; 0.487)

**Tabela 5.3:** Ilustração comparativa dos algoritmos - Medidas de distância e redução percentual relativa para os modelos *Logístico*

Modelo	Cramér Von-Mises	Kolmogorov-Smirnov	Anderson-Darling
BP	0.00304	0.08769	0.01543
BSP ME	0.00306	0.08431	0.01528
BSP MEVL	0.00309	0.08358	0.0154

As estimativas pontuais dos modelos BP e BSP MEVL são similares, embora, os intervalos de credibilidade 95% mais extensos para o último modelo indiquem que há mais uma componente responsável pela variabilidade sendo estimada neste modelo, o que é esperado. Em contrapartida,

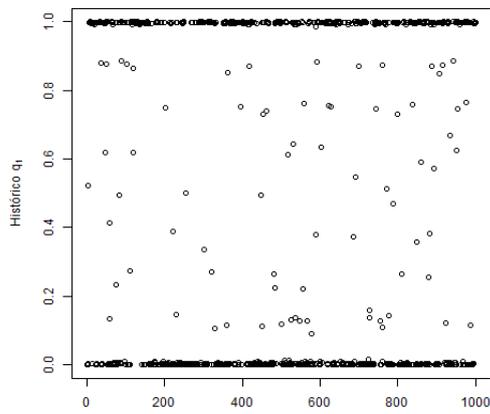
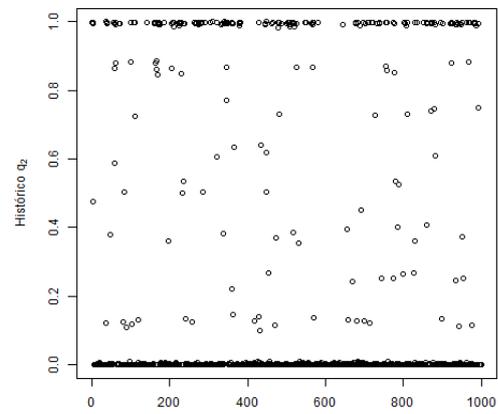
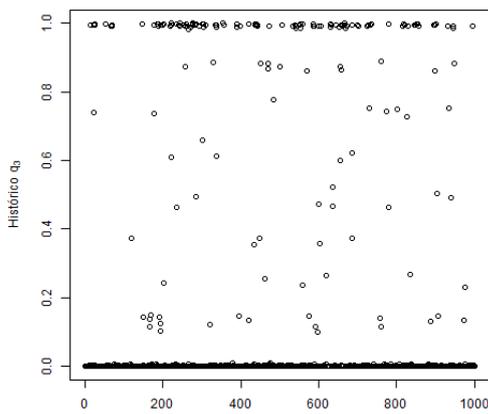
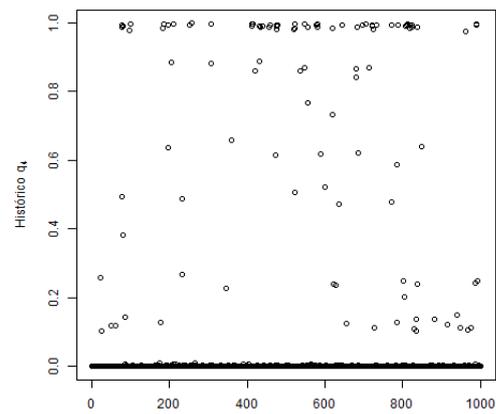
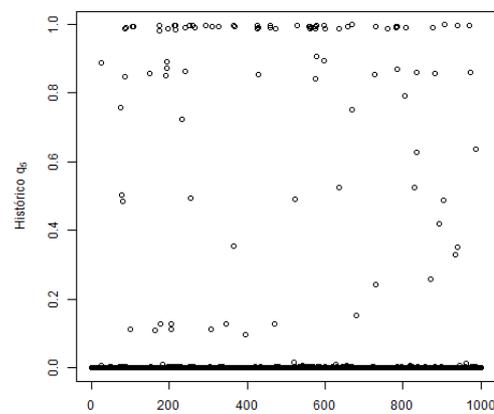
as estimativas do modelo BSP ME são discrepantes em relação ao modelo BP, ainda que as medidas de distância sejam bastante similares entre todos os modelos.

Pode-se visualizar as cadeias de Markov que definem os pesos das componentes do processo *Dirichlet* nas Figuras 5.1 e 5.2. De fato, como discutido por Ferguson [1983], para um pequeno valor de  $\alpha$  segue que a primeira componente do processo *Dirichlet* tem maior influência na mistura.

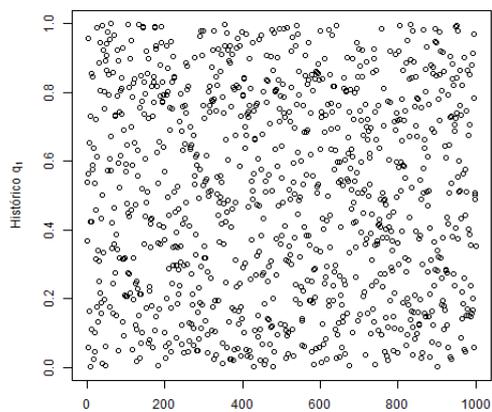
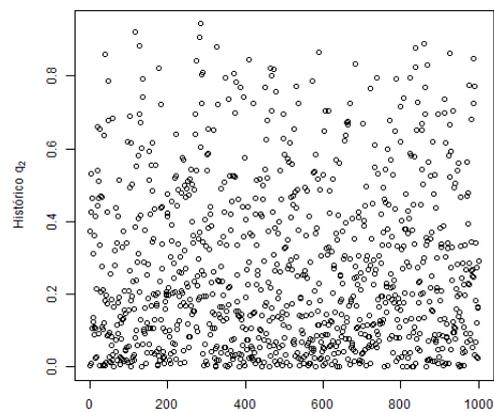
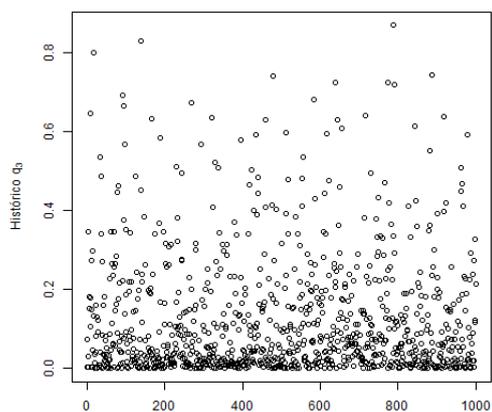
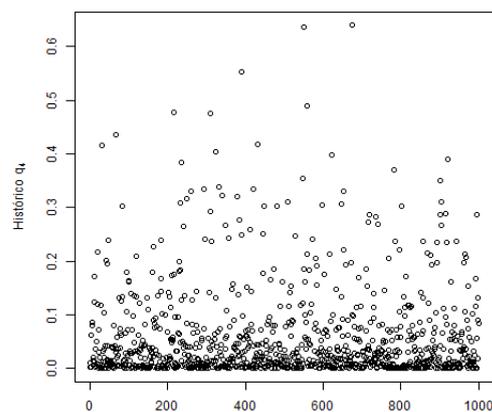
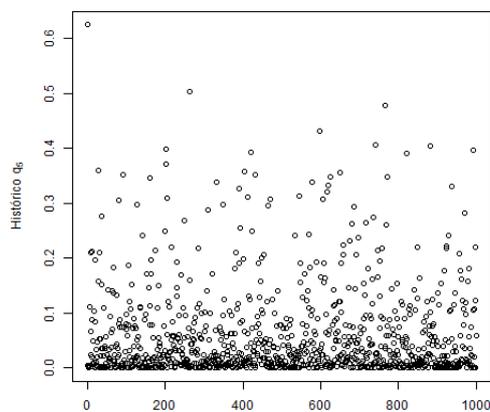
Os dois conjuntos de gráficos permitem verificar que o vetor de pesos  $\mathbf{q}$  apresenta um comportamento distinto para os dois algoritmos. Na abordagem através de variáveis latentes, em geral, cada componente tem peso um em cada interação, o que não caracteriza uma mistura de distribuições e é bastante diferente da abordagem sem variáveis latentes.

As diferenças nas estimativas na Tabela 5.2 podem ser justificada através de tais percepções, uma vez que a soma de distribuições *Logística* não resulta em uma distribuição *Logística*, por isso, é esperado que a estimativa dos coeficientes de regressão  $\beta$  no modelo semi-paramétrico de mistura em escala seja distinta do modelo paramétrico.

Portanto, tais indícios levam a crer que a estimação do modelo bayesiano semi-paramétrico de mistura em escala através de variáveis latentes, na verdade, corresponde à estimação da variância do modelo *Logístico* paramétrico, ao invés de uma mistura de distribuições *Logística*.

(a) *Histórico  $q_1$* (b) *Histórico  $q_2$* (c) *Histórico  $q_3$* (d) *Histórico  $q_4$* (e) *Histórico  $q_5$* 

**Figura 5.1:** *Histórico das cadeias de Markov para os pesos do processo Dirichlet do modelo MEVL*

(a) *Histórico  $q_1$* (b) *Histórico  $q_2$* (c) *Histórico  $q_3$* (d) *Histórico  $q_4$* (e) *Histórico  $q_5$* **Figura 5.2:** *Histórico das cadeias de Markov para os pesos do processo Dirichlet do modelo ME*

## Capítulo 6

# Considerações Finais

Neste trabalho, um novo modelo *Logístico* bayesiano paramétrico com um parâmetro representativo da assimetria e três novos modelos *Logístico* bayesianos semi-paramétricos, com diferentes níveis de sofisticação para lidar com discrepâncias da distribuição *Logística*, foram apresentados. Para avaliar o desempenho de tais modelos, três medidas de distância entre medidas de probabilidade foram sugeridas, além das Ordenadas Preditivas Condicionais.

Em todos os modelos, é possível elicitar a distribuição a priori dos parâmetros através do conceito de razão de chance, por isso, a modelagem sugerida se sobrepõe à modelagem apresentada por [Newton \*et al.\* \[1996\]](#) na aplicação prática. Contudo, a interpretação dos coeficientes de regressão a posteriori não é direta, embora ainda seja possível discutir as razões de chances uma vez que a curva de probabilidade está estimada.

No Capítulo 3, três exemplos foram considerados. No exemplo simulado simétrico, todos os modelos apresentaram um desempenho similar em relação às três medidas de distância consideradas, com um pequeno destaque para o modelo bayesiano semi-paramétrico com mistura em posição e escala, sendo que o modelo bayesiano paramétrico usual tem o apelo da interpretação a posteriori dos parâmetros através do conceito de razão de chances.

Na presença de assimetria, os modelos bayesianos paramétrico e semi-paramétrico com transformação apresentaram desempenhos similares, acompanhados da caracterização da assimetria através de um parâmetro em relação ao modelo bayesiano semi-paramétrico com mistura em posição e escala que apresentou desempenho mediano, porém, satisfatório.

Na presença de bimodalidade, o modelo semi-paramétrico de posição e escala destaca-se entre todos os modelos para todas as medidas de distâncias. Este é o modelo mais geral, apesar de não ser o mais interessante em situações mais simples, a sua generalidade é bastante útil frente a situações de total ignorância sobre a forma da curva de probabilidade.

Ainda é necessário realizar simulações para que seja possível avaliar o desempenho dos modelos frente a estas situações.

Para os dois conjuntos de dados reais no Capítulo 4, os modelos semi-paramétricos e paramétrico com transformação são competitivos com os modelos mais apropriados apresentados até o momento na literatura estatística. Para a mortalidade dos besouros, a vantagem da modelagem introduzida é a presença de um parâmetro que mensura a assimetria frente ao modelo *Log-log Complementar*, enquanto, para o conjunto de dados Idade da primeira menstruação, a vantagem é a simplicidade diante do modelo de [Stukel \[1988\]](#).

A grande limitação deste trabalho é a implementação computacional. No capítulo 5, foi visto que a abordagem de variáveis latentes introduzida por [Albert e Chib \[1993\]](#) parece não ser adequada no contexto bayesiano não-paramétrico por não possibilitar a construção de misturas. Por outro lado, a utilização do algoritmo Metropolis-Hastings Adaptativo é bastante custosa computacionalmente e torna-se um empecilho na aplicação cotidiana de tais modelos. Assim, um trabalho futuro seria a busca por métodos computacionais mais eficientes, possivelmente, uma nova representação ou um novo conjunto de condições para variáveis latentes, assim como [Polson \*et al.\* \[2013\]](#).

Além disso, cabe ressaltar que o modelo bayesiano paramétrico com transformação apresentou

desempenho marcante em quase todas os exemplos estudados, e como a abordagem de variáveis latentes é funcional no contexto paramétrico, um algoritmo pode ser implementado similarmente a [Czado \[1994\]](#).

E para verificar a hipótese pontual de que o modelo simétrico é adequado, pode-se implementar o teste FBST apresentado por [Pereira e Stern \[1999\]](#), pois, a utilização dos intervalos de credibilidade 95% pode não ser adequada frente a distribuições a posteriori multivariadas. Além disso, este modelo também pode ser explorado no contexto frequentista. Por fim, tais modelos podem ser facilmente expandidos para dados multinomiais, através da reparametrização introduzida por [Pereira e Stern \[2008\]](#).

# Referências Bibliográficas

- Albert e Chib (1993)** James H. Albert e Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679. Citado na pág. 2, 5, 6, 37, 41, 45, 49
- Anderson e Darling (1952)** Ted W Anderson e Donald A Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, páginas 193–212. Citado na pág. 18
- Andrews e Mallows (1974)** David F. Andrews e Colin L Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 99–102. Citado na pág. 9
- Antoniak (1974)** Charles E. Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, páginas 1152–1174. Citado na pág. 7, 15
- Aranda-Ordaz (1981)** Francisco J. Aranda-Ordaz. On two families of transformations to additivity for binary response data. *Biometrika*, 68(2):357–363. Citado na pág. 2, 11
- Azzalini (1985)** Adelchi Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 12:171–178. Citado na pág. 2
- Basu e Mukhopadhyay (1998)** S. Basu e S. Mukhopadhyay. Binary response regression with normal scale mixture links. Em Ghosh Sujit K Dey, Dipak K e Bani K Mallick, editors, *Generalized linear models: A Bayesian perspective*, páginas 231–241. Marcel Dekker, New York. Citado na pág. iii, v, 2, 3, 9
- Bazan et al. (2010)** Jorge L. Bazan, Heleno Bolfarine e Marcia D. Branco. A framework for skew-probit links in binary regression. *Communications in Statistics-Theory and Methods*, 39(4): 678–697. Citado na pág. 2
- Bickel e Doksum (1981)** Peter J. Bickel e Kjell A. Doksum. An analysis of transformations revisited. *Journal of the american statistical association*, 76(374):296–311. Citado na pág. 11
- Blackwell e MacQueen (1973)** David Blackwell e James B. MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, páginas 353–355. Citado na pág. 7
- Bliss (1934)** Chester I. Bliss. The method of probits. *Science*, 79(2037):38–39. Citado na pág. 1
- Bliss (1935)** Chester I. Bliss. The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22(1):134–167. Citado na pág. 1, 5, 29
- Box e Cox (1964)** George E. P. Box e David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 211–252. Citado na pág. iii, v, 2, 3, 11
- Burnham e Anderson (2002)** Kenneth P. Burnham e David R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media. Citado na pág. 6

- Caron e Polpo (2009)** Renault Caron e Adriano Polpo. Binary data regression: Weibull distribution. Em *Bayesian Inference and Maximum Entropy methods in Science and Engineering: The 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1193, páginas 187–193. AIP Publishing. Citado na pág. 2
- Chen e Dey (1998)** Ming-Hui Chen e Dipak K. Dey. Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhyā: The Indian Journal of Statistics, Series A*, páginas 322–343. Citado na pág. 2, 6, 10, 40, 41
- Chen et al. (1999)** Ming-Hui Chen, Dipak K. Dey e Qi-Man Shao. A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94(448): 1172–1186. Citado na pág. 2
- Connor e Mosimann (1969)** Robert J. Connor e James E. Mosimann. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, 64(325):194–206. Citado na pág. 8
- Cox (1969)** David R. Cox. *The Analysis of Binary Data*. Chapman and Hall, London. Citado na pág. 1, 5
- Cramer (1928)** Harald Cramer. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74. Citado na pág. 18
- Cramer (2011)** Jan S. Cramer. *Logit models from economics and other fields*. Cambridge University Press. Citado na pág. 1
- Czado (1994)** Claudia Czado. Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference*, 41(2):121–140. Citado na pág. 2, 11, 50
- Devroye (1986)** L. Devroye. *Non-uniform Random Variate Generation*. Springer-Verlag. Citado na pág. 10
- Escobar (1994)** Michael D. Escobar. Estimating normal means with a dirichlet process prior. *Journal of the American Statistical Association*, 89(425):268–277. Citado na pág. 2, 15, 38
- Escobar e West (1995)** Michael D. Escobar e Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588. Citado na pág. 38
- Escobar e West (1998a)** Michael D. Escobar e Mike West. Computing nonparametric hierarchical models. Em Dipak D. Dey, Peter Muller e Debajyoti Sinha, editors, *Practical Nonparametric and Semiparametric Bayesian Statistics*, páginas 1–22. Springer, New York. Citado na pág. 38
- Escobar e West (1998b)** Michael D. Escobar e Mike West. Computational methods for mixture of dirichlet process models. Em Dipak D. Dey, Peter Müller e Debajyoti Sinha, editors, *Practical Nonparametric and Semiparametric Bayesian Statistics*, páginas 23–43. Springer, New York. Citado na pág. 38
- Ferguson (1973)** Thomas S. Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, páginas 209–230. Citado na pág. 6
- Ferguson (1983)** Thomas S. Ferguson. Bayesian density estimation by mixtures of normal distributions. *Recent advances in statistics*, 24:287–302. Citado na pág. 13, 46
- Gamerman (1997)** Dani Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68. Citado na pág. 6
- Gelfand (1996)** Alan E Gelfand. Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, páginas 145–161. Citado na pág. 6

- Guerrero e Johnson (1982)** Victor M. Guerrero e Richard A. Johnson. Use of the box-cox transformation with binary response models. *Biometrika*, 69(2):309–314. Citado na pág. [iii](#), [v](#), [2](#), [3](#), [11](#)
- Ishwaran e James (2001)** Hemant Ishwaran e Lancelot F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453). Citado na pág. [8](#), [15](#), [37](#), [38](#), [39](#)
- Ishwaran e Zarepour (2000)** Hemant Ishwaran e Mahmoud Zarepour. Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371–390. Citado na pág. [8](#), [38](#)
- John e Draper (1980)** J. A. John e N. R. Draper. An alternative family of transformations. *Applied Statistics*, páginas 190–197. Citado na pág. [11](#)
- Kolmogorov (1933)** Andrej N Kolmogorov. *Sulla determinazione empirica di una legge di distribuzione*, volume 4. Citado na pág. [18](#)
- Lo (1984)** Albert Y Lo. On a class of bayesian nonparametric estimates: I. density estimates. *The annals of statistics*, 12(1):351–357. Citado na pág. [12](#)
- MacEachern (1994)** Steven N. MacEachern. Estimating normal means with a conjugate style dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3):727–741. Citado na pág. [38](#)
- MacEachern e Müller (1998)** Steven N. MacEachern e Peter Müller. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2):223–238. Citado na pág. [38](#)
- Malthus (1872)** Thomas R. Malthus. *An Essay on the Principle of Population Or a View of Its Past and Present Effects on Human Happiness, an Inquiry Into Our Prospects Respecting the Future Removal Or Mitigation of the Evils which it Occasions by Rev. TR Malthus*. Reeves and Turner. Citado na pág. [1](#)
- Manly (1976)** B. F. J. Manly. Exponential data transformations. *The Statistician*, páginas 37–42. Citado na pág. [2](#), [11](#), [12](#)
- Milicer et al. (1968)** Halina Milicer, e Franciszek Szczotka. Age at menarche in warsaw girls in 1965. *Human biology*, páginas 249–259. Citado na pág. [33](#)
- Nelder e Weddeburn (1972)** J. A. Nelder e R. W. M. Weddeburn. Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384. Citado na pág. [1](#)
- Newton et al. (1996)** Michael A. Newton, Claudia Czado e Rick Chappell. Bayesian inference for semiparametric binary regression. *Journal of the American Statistical Association*, 91(433):142–153. Citado na pág. [iii](#), [v](#), [2](#), [3](#), [49](#)
- Pereira e Stern (1999)** Carlos Alberto de Bragança Pereira e Julio Michael Stern. Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110. Citado na pág. [50](#)
- Pereira e Stern (2008)** Carlos Alberto de Bragança Pereira e Julio Michael Stern. Special characterizations of standard discrete models. *REVSTAT-Statistical Journal*, 6(3):199–230. Citado na pág. [50](#)
- Pitman e Yor (1997)** Jim Pitman e Marc Yor. The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900. Citado na pág. [8](#)

- Polson et al. (2013)** Nicholas G Polson, James G Scott e Jesse Windle. Bayesian inference for logistic models using pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349. Citado na pág. [2](#), [49](#)
- Prentice (1976)** Ross L. Prentice. A generalization of the probit and logit methods for dose response curves. *Biometrics*, páginas 761–768. Citado na pág. [2](#)
- Robert (1995)** Christian P. Robert. Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125. Citado na pág. [44](#)
- Sahu et al. (2003)** Sujit K. Sahu, Dipak K. Dey e Marcia D. Branco. A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics*, 31(2):129–150. Citado na pág. [2](#)
- Sakia (1992)** R. M. Sakia. The box-cox transformation technique: a review. *The statistician*, páginas 169–178. Citado na pág. [10](#)
- Schlaifer e Raffa (1961)** Robert Schlaifer e Howard Raffa. *Applied Statistical Decision Theory*. Wiley-Interscience, New York, 1ª ed. Citado na pág. [6](#)
- Sethuraman (1994)** Jayaram Sethuraman. A constructive defintion of dirichelt priors. *Statistica Sinica*, 4:639–650. Citado na pág. [7](#)
- Smirnov (1948)** Nickolay Smirnov. Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, páginas 279–281. Citado na pág. [18](#)
- Spiegelhalter et al. (2002)** David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin e Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639. Citado na pág. [6](#)
- Stukel (1988)** Thérèse A. Stukel. Generalized logistic models. *Journal of the American Statistical Association*, 83(402):426–431. Citado na pág. [iii](#), [v](#), [2](#), [3](#), [11](#), [35](#), [49](#)
- Tukey (1957)** John W. Tukey. On the comparative anatomy of transformations. *The Annals of Mathematical Statistics*, páginas 602–632. Citado na pág. [10](#)
- von Mises (1928)** R von Mises. *Wahrscheinlichkeit, Statistik and Wahrheit*. Citado na pág. [18](#)
- Wong (1998)** Tzu-Tsung Wong. Generalized dirichlet distribution in bayesian analysis. *Applied Mathematics and Computation*, 97(2):165–181. Citado na pág. [43](#)
- Yeo e Johnson (2000)** In-Kwon Yeo e Richard A. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959. Citado na pág. [iii](#), [v](#), [3](#), [11](#), [12](#)
- Zwet (1964)** Willem Rutger Zwet. *Convex transformations of random variables*. Mathematisch centrum. Citado na pág. [11](#)

Bibliografia