

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Análises Bayesianas para o modelo de Regressão
Birnbaum-Saunders com zeros ajustados**

Jadson Luan dos Santos Marcelino

Dissertação de Mestrado do Programa Interinstitucional de
Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Jadson Luan dos Santos Marcelino

Análises Bayesiana para o modelo de Regressão Birnbaum-Saunders com zeros ajustados

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *EXEMPLAR DE DEFESA*

Área de Concentração: Estatística

Orientadora: Profa. Dra. Vera Lucia
Damasceno Tomazella

USP – São Carlos
Agosto de 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

d722a dos Santos Marcelino, Jadson Luan
Análise Bayesiana para o modelo de regressão
Birnbau-Saunders com Zeros Ajustados / Jadson Luan
dos Santos Marcelino; orientadora Vera Lúcia
Damsceno Tomazella. -- São Carlos, 2021.
111 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2021.

1. MODELAGEM DE DADOS. 2. DISTRIBUIÇÕES
(PROBABILIDADE). I. Damsceno Tomazella, Vera Lúcia
, orient. II. Título.

Jadson Luan dos Santos Marcelino

**Bayesian analysis for Birnbaum-Saunders regression
model with adjusted zeros**

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP
and to the Department of Statistics – DEs-UFSCar, in
partial fulfillment of the requirements for the degree of
the Master Interagency Program Graduate in Statistics.
EXAMINATION BOARD PRESENTATION COPY

Concentration Area: Statistics

Advisor: Profa. Dra. Vera Lucia
Damasceno Tomazella

USP – São Carlos
August 2021



UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Jadson Luan dos Santos Marcelino, realizada em 11/08/2021.

Comissão Julgadora:

Profa. Dra. Vera Lucia Damasceno Tomazella (UFSCar)

Prof. Dr. Manoel Ferreira dos Santos Neto (UFMG)

Prof. Dr. Juvêncio Santos Nobre (UFPA)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

*Este trabalho é dedicado ao meu amor **André Yoshida Machado** que sempre me acolheu tanto, me incentivou e até me ajudou no ingresso no mestrado. Você é e sempre será muito precioso para mim.*

AGRADECIMENTOS

Nesta sessão gostaria de fazer agradecimento de maneira diferente, aqui gostaria de contar um pouco sobre como cheguei até aqui e justificar assim os agradecimentos que gostaria de fazer.

Sou Pernambucano, nascido em Ipojuca (cidade onde fica localizada a praia de porto de galinhas). Como este pernambucano veio parar aqui? Pois bem, enquanto vivia em Pernambuco, com família bastante simples e configuração familiar complicada, era necessário que desde muito pequeno começasse a trabalhar. Então, o tempo mínimo que “esperei” para trabalhar foi o mesmo que tive para aprender a fazer contas de subtração, adição, divisão e multiplicação. Isso por que era necessário saber estas operações básicas para conseguir dar troco enquanto vendia toda sorte de alimento vinda dos rios, roça ou mares. Meu pai, Antônio Fracisco Marcelino, é analfabeto, assim como minha mãe (Maria José dos Santos). Ele saía de casa todos os dias para ir ao mangue capturar carangueijos, camarões, sirí, unha, marisco, etc. Ou mesmo quando fazia mais sentido ir à roça era de lá que vinham os alimentos para vender e trazer algum dinheiro para casa bem como para complementar a alimentação. Foi assim que comecei minha vida de trabalho, saindo pelas ruas de Nossa Senhora do Ó, gritando aos moradores sobre o que estava vendendo.

Quando mais velho, fui também para as lindas casas de veraneio de Porto de Galinhas para construção de piscinas, construção das saunas, pintura de lindos quartos, etc. E assim conseguia, junto com um dos meus irmãos, trazer algum dinheiro adicional (e essencial) para casa.

Quando um tanto mais velho, lá com meus 11 anos, mudei de ramo e passei a catar latas também em Porto de Galinhas com o mesmo objetivo: trazer dinheiro para casa e ajudar a situação familiar.

Foi aí então que, em uma dessas idas e vindas de catação de latas, eu conheci a

Wânia Stocche Barbosa e parte de sua família. Por sorte do destino, subi os degraus da casa onde estavam em busca de latas para poder vender. Ela então fez muito mais que isso. Nos convidou para subir, sentar e almoçar antes mesmo de entregar as latas que tanto esperávamos.

Ao receber estas latas, me impressionei pela quantidade que tinha. Como estavam em muitos na casa, a quantidade era bastante boa: O suficiente para não mais precisar andar de ponta a ponta da praia vasculhando os lixos para então poder encher um saco inteiro de latas. Sabendo disso, então, mudamos nossa estratégia e ao invés de passar por todos as lixeiras íamos diretamente para essa casa onde tínhamos certeza da existência de latas, e em grande quantidade.

Foi assim então que me aproximei da Wânia, que em uma dessas idas minhas me questionou se gostava de estudar. Para mim, estudar nem era uma questão de gosto, pois era na escola onde me sentia e me permitia ser criança, além de tantas outras vantagens que tinha por estar dentro das salas de aula, muitas delas nem diziam respeito ao conteúdo em si que era passado.

Wânia se solidarizou e conheceu minha casa, que na época num passava de um barraco construído com plásticos pois a casa de fato, que era feita de barro e madeiras retiradas do mangue, precisou ser derrubada por conta dos cupins que estavam comendo as madeiras e criando algum risco de queda. Para que isso não acontecesse enquanto estávamos dentro da casa, preferimos derrubar para poder construir outra do zero da mesma maneira. Mas desta vez foi diferente. Wânia, antes de voltar para sua cidade de origem, comprou todos os materiais para construção de uma casa de alvenaria, também comprou sofá, geladeira, uma bicicleta que me ajudaria a ir a escola e tanto mais. Com uma única condição: que eu parasse de trabalhar e estudasse integralmente em uma escola que ela começaria a pagar.

Foi aí que minha vida de fato começou a mudar.

Depois de pagar meus estudos desde a sétima série até o segundo colegial, aos meus 17 anos Wânia me convida para finalizar estes estudo em Ribeirão Preto (cidade onde vive) para assim poder entrar em uma universidade e cursar o ensino superior. E assim foi! Enquanto trabalhava durante o dia e estudava durante a noite, já em Ribeirão Preto, aos finais de semana me dedicava “tirar o atraso” daquilo que foi dado em aula

naquela semana. Com isso, com muito custo, entrei na UFSCar no curso de Estatística. Com pouco conhecimento de matemática, mas com muito amor por ela.

E aqui deixo o meu maior agradecimento a alguém que de fato modificou minha vida e possibilitou ter a que tenho agora. Wânia Stocche Barbosa. Muito muito muito obrigado. A minha gratidão, a minha maneira, será eterna.

Na faculdade não foi diferente, tive diversos auxílios. Do governo da época que facilitava o acesso ao ensino superior de pessoas carentes por meio de concessão de bolsas: moradia, alimentação e uma bolsa de estudo.

Morei no alojamento da faculdade, comia TODOS, sem exceção, os dias no restaurante universitário, pois comia de graça e com o dinheiro que recebia por trabalhar na biblioteca do departamento (Um agradecimento a Profa. Dra. Cristina) podia de vez em quando me juntar aos colegas em uma pizzaria ou barzinho para desafogar o estresse do dia a dia de um curso que não é fácil.

Por sorte minha, e olha quanta sorte, viu? Minha primeira professora foi a que hoje me orienta na condução deste trabalho e que também me conduziu no trabalho de conclusão de curso, a professora Dra. Vera Tomazella (que com carinho chamo de Verinha). É alguém que nunca soltou minha mão, mesmo quando ela tinha certeza que eu deveria seguir na carreira acadêmica e eu hesitava por conta da necessidade de ganhar dinheiro rapidamente (e possibilitar uma vida diferente aos meus), ela esteve alí do meu lado, entendeu a dificuldade adicional que tive ao começar a trabalhar, mesmo ainda precisando finalizar algumas disciplinas e também o TCC (aqui ela puxava bastante minha orelha pra fazer o meu melhor sempre). E assim o foi. Depois de tempos no mercado de trabalho, e também por diversas dificuldades que encontrei na vida adulta (e não foram poucas), decidi retornar e dar ouvidos ao que esta mulher maravilhosa há muito já havia me falado. Decidi então voltar a academia para o mestrado e não poderia ter outra pessoa me orientado que não a Profa. Dra. Vera.

Por isso, é extremamente necessário que ela tenha esse espaço, que pra mim é especial, aqui na minha dissertação (dissertação essa que é mais nossa que minha). Sem ela, mais uma vez, não chegaria a esta etapa. Ela sabe quão difícil foi e também sabe quão importante foi nessa construção. Segurou minha mão diversas vezes quando me via incapacitado para continuar.

Não somente ela, também o Gustavo Sabillon, a Camila Ozelame e a Débora Stern foram essenciais em todo esse processo do mestrado. Estes últimos, desde o primeiro dia de aula do processo seletivo estiveram juntos. E quando digo juntos é juntos mesmo. Saíamos da aula de Teoria das probabilidades e seguíamos para a lousa do ICMC onde refazíamos tudo que havíamos aprendido naquele dia a fim de conseguir acompanhar a matéria e passar bem para termos a bolsa que possibilitava nossos estudos. E, sem sombra de dúvida, sem eles eu não teria conseguido. Sem o apoio que um dava ao outro a cada exercício que nos deparávamos, sem a presença deles à meia noite estudando e quebrando a cabeça para entender aqueles conceitos (muitos tão abstratos, num é mesmo sigma-álgebra?). Mas conseguimos e eu só consegui por vocês estarem sempre presente e dispostos. Muita gratidão a vocês. Minha grande chiminéia.

Agora no final deste processo longo, preciso também agradecer outros tantos que contribuíram e muito em todo esse desenvolvimento: o Prof. Dr. Manoel Santos-Neto, o Pesquisador Dr. Márcio Diniz, e todos aqueles que tanto me apoiaram até aqui.

Difícil listar nomes sem esquecer de alguém e neste último parágrafo agradeço a todos que de alguma maneira me auxiliaram a chegar até aqui. Muita muita gradidão por tudo.

RESUMO

MARCELINO, J. L. S. **Análises Bayesianas para o modelo de Regressão Birnbaum-Saunders com zeros ajustados**. 2021. 108 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

O modelo probabilístico Birnbaum-Saunders vem recebendo atenção considerável nos últimos anos. Neste trabalho nós consideramos a distribuição Birnbaum-Saunders reparametrizada com zeros ajustados (ZARBS) (SANTOS-NETO *et al.*, 2012). Esta distribuição admite a ocorrência de zeros com probabilidade positiva, considerando um modelo de mistura discreta-contínua que é construído usando uma massa de probabilidade no zero e uma componente contínua. A ZARBS generaliza pelo menos sete modelos de regressão Birnbaum-Saunders reparametrizados. Neste contexto a principal contribuição desta dissertação é estudar a ZARBS sob enfoque Bayesiano utilizando o pacote BAMLSS desenvolvido no software R, bem como derivar diagnósticos de influência para o modelo. Os métodos de diagnóstico têm sido ferramentas importantes na análise de regressão para detectar anomalias, tais como quebras das pressuposições na parte estocástica do modelo, presença de outliers e observações influentes. Nós avaliamos a influência local nas estimativas dos parâmetros considerando um esquema de perturbação. Para verificar as potencialidades da metodologia proposta consideramos uma aplicação de um conjunto de dados reais.

Palavras-chave: Birnbaum-Saunders, Reparametrização, Zeros Ajustados, Estimação, Análise Bayesiana.

ABSTRACT

MARCELINO, J. L. S. **Bayesian analysis for Birnbaum-Saunders regression model with adjusted zeros**. 2021. 108 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Modeling based on the Birnbaum-Saunders distribution has received considerable attention in recent years. In this work we consider the reparametrized Birnbaum-Saunders distribution with zero-adjusted (ZARBS) ([SANTOS-NETO *et al.*, 2012](#)). This distribution admits the occurrence of zeros with positive probability, considering a discrete-continuous mixing model that is constructed using a probability mass at zero and a continuous component. ZARBS generalizes at least seven reparametrized Birnbaum-Saunders regression models. In this context, the main contribution of this dissertation is to study ZARBS under a Bayesian approach using the BAMLSS package developed in the R software, as well as to derive influence diagnoses for the model. Diagnostic methods have been important tools in regression analysis to detect anomalies, such as breaking assumptions in the stochastic part of the model, presence of outliers and influential observations. We assess the local influence on parameter estimates considering a perturbation scheme. To verify the potential of the proposed methodology, an application to a set of real data will be considered.

Keywords: Birnbaum-Saunders Distribution, Reparameterized, Adjusted zeros, Estimation, Bayesian Analysis.

LISTA DE ILUSTRAÇÕES

Figura 2.1.1-Função de densidade da BS para diferentes parâmetros.	29
Figura 2.1.2-Função de Distribuição para diferentes parâmetros.	30
Figura 2.2.1-Função Densidade de Probabilidade da variável aleatória RBS para diferentes valores dos parâmetros	32
Figura 3.2.1-Funcionamento do pacote.	57
Figura 3.2.2-Comparação entre os participantes da força de trabalho e os não participantes variando a idade, curva azul (extrangeiros) e curva preta (não estrangeiros).	63
Figura 4.3.1-Gráfico de Evelope Simulado.	77
Figura 4.3.2-Gráfico de Dispersão dos Resíduos pelos Índices.	77
Figura 4.3.3-Gráficos para Análise de Pontos Influentes.	78
Figura 5.3.1-Medida KL para o modelo final ajustado.	88
Figura A.0.1-Gráficos para checagem da convergência para os parâmetros estima- dos.	100
Figura A.0.2-Gráficos para checagem da convergência para os parâmetros estima- dos.	101
Figura A.0.3-Gráficos para checagem da convergência para os parâmetros estima- dos.	102
Figura A.0.4-Gráficos para checagem da convergência para os parâmetros estima- dos.	103
Figura B.0.1-Gráficos para checagem da convergência para os parâmetros estima- dos.	106
Figura B.0.2-Gráficos para checagem da convergência para os parâmetros estima- dos.	107
Figura B.0.3-Gráficos para checagem da convergência para os parâmetros estima- dos.	108

LISTA DE TABELAS

Tabela 1 – Estimativas dos Parâmetros para o modelo ZARBS.	74
Tabela 2 – Estimativas dos Parâmetros para o modelo ZARBS.	74
Tabela 3 – Estimativas dos Parâmetros para o modelo ZARBS.	75
Tabela 4 – Estimativas dos Parâmetros para o modelo ZAIG.	75
Tabela 5 – Estimativas dos Parâmetros para o modelo ZAGA.	76
Tabela 6 – Medidas AIC e SBC para os modelos ajustados	76
Tabela 7 – Estimativas dos coeficientes obtidas usando o pacote bamlss.	85
Tabela 8 – Estimativas dos coeficientes obtidas usando o pacote bamlss.	86
Tabela 9 – Comparação das estimativas dos parâmetros na abordagem Clássica vs Bayesiana	87

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivo	24
1.2	Organização do Trabalho	25
2	PRELIMINARES	27
2.1	A Distribuição Birnbaum-Saunders (BS)	27
2.1.1	<i>Funções Densidade e Distribuição de Probabilidade</i>	28
2.2	Distribuição Birnbaum-Saunders Reparametrizada	31
2.3	Inferência Bayesiana	33
2.3.1	<i>Distribuições a Priori</i>	35
2.3.1.1	<i>Conjugação na Família Exponencial</i>	37
2.3.1.2	<i>Priori não-informativa</i>	39
2.3.2	<i>Critérios de seleção de modelos</i>	44
2.3.3	<i>Conclusão</i>	46
3	FERRAMENTA COMPUTACIONAL	47
3.1	Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS)	48
3.2	Pacote BAMLSS	51
3.2.1	<i>Resposta e distribuição posteriori</i>	53
3.2.2	<i>Ajuste do modelo</i>	54
3.2.3	<i>Lego Bricks</i>	56
3.2.4	<i>Exemplo de uso do pacote</i>	58
4	O MODELO DE REGRESSÃO ZARBS	65
4.1	Modelo ZARBS	65
4.1.1	<i>Vetor Score, Matriz Hessiana e de Informação de Fisher</i>	67

4.2	Análise de Diagnóstico	70
4.2.1	<i>Resíduos Quantílicos</i>	70
4.2.2	<i>Influência Local</i>	71
4.3	Exemplo de Aplicação	72
4.3.1	<i>Resultados Abordagem Clássica</i>	74
4.3.2	<i>Análise de Diagnóstico</i>	76
5	UMA ABORDAGEM BAYESIANA PARA O MODELO DE REGRESSÃO ZARBS	81
5.1	O BAMLSS para o modelo de regressão ZARBS	82
5.2	Distribuição a Posteriori	82
5.3	Aplicação	84
5.3.1	<i>Diagnóstico Bayesiano</i>	87
6	CONCLUSÕES PRELIMINARES E PROPOSTA FUTURA .	91
	REFERÊNCIAS	93
APÊNDICE A	GRÁFICOS DE CONVERGÊNCIA PARA O MO- DELO COM TODAS VARIÁVEIS	99
APÊNDICE B	GRÁFICOS DE CONVERGÊNCIA PARA O MO- DELO FINAL	105

INTRODUÇÃO

Um dos grandes problemas na área da indústria é quando acontece a ruptura ou falha dos materiais. Estas falhas são devidas a diversos fatores como por exemplo, a fadiga dos materiais. Designa-se por fadiga uma forma de falha de materiais sujeitos a ciclos repetidos de tensão ou deformação. Este fenômeno é verificado principalmente em aeronaves, componentes de máquinas, pontes, entre outros elementos mecânicos. Modelos estatísticos para processos de fadiga (dano estrutural ocorrido quando um material é exposto a alguma tensão ou esforço) permitem descrever a variação aleatória dos tempos de falha. Alguns modelos probabilísticos têm sido propostos para descrever a vida por fadiga, dentre eles, encontram-se as distribuições Gama, Normal Inversa, Lognormal e Weibull. A distribuição Birnbaum-Saunders (BS) foi proposta por [Birnbaum e Saunders \(1969a\)](#) e está baseada em um argumento físico de dano cumulativo que produz a fadiga de materiais. E é neste cenário que a distribuição foi criada, tirando daí a ideia de que equipamentos submetidos a vários ciclos de cargas podem ter suas funções danificadas, ver [Birnbaum e Saunders \(1969a\)](#).

[Mann, Schafer e Singpurwalla \(1974\)](#) mostraram que a densidade da BS possui apenas uma moda, ou seja, unimodal. [Engelhardt, Bain e Wright \(1981\)](#) obtiveram intervalos de confiança e teste de hipóteses para os parâmetros da distribuição, [Rieck e Nedelman \(1991\)](#) derivou a função geradora de momentos para a distribuição Senh-normal que pode ser usada para obter momentos de ordem inteira ou fracionária da

distribuição BS. [Achcar \(1993\)](#) desenvolveu procedimentos de estimação bayesiana para os parâmetros usando aproximações para as distribuições posterioris marginais. [Birnbaum e Saunders \(1969b\)](#) determinaram originalmente os estimadores de máxima verossimilhança para os parâmetros do modelo, [Lu e Chang \(1997\)](#) utilizaram métodos bootstrap para construir intervalos de predição para a distribuição BS. [Dupuis e Mills \(1998\)](#) propuseram métodos robustos de estimação dos parâmetros da distribuição BS. [Galea, Leiva-Sánchez e Paula \(2004\)](#) fizeram uma análise de diagnóstico no modelo log-Birnbaum-Saunders entre muito outros. Autores como [Leiva et al. \(2008a\)](#), [Leiva et al. \(2008b\)](#), [Leiva et al. \(2008c\)](#). [Kundu, Balakrishnan e Jamalizadeh \(2013\)](#) obtiveram estimadores por via do método de momentos modificado.

Generalizações da BS foram determinadas em diversas vertentes. Por exemplo, em alguns estudos foram considerados novos parâmetros com a finalidade de se obter modelos mais robustos e flexíveis de tal forma que fossem aplicados em problemas reais. [Owen e Padgett \(1999\)](#) incluíram um terceiro parâmetro à distribuição BS. [Díaz-García e Leiva \(2005\)](#) foram adiante e desenvolveu uma generalização sob a classe das distribuições elípticas que são mais flexíveis quanto a curtose e assimetria. [Xie e Wei \(2007\)](#) estudaram medidas de diagnóstico para modelo de regressão log-BS. Em [Ahmed et al. \(2008\)](#) foi apresentada uma nova reparametrização que ajusta fenômenos físicos uma vez que os parâmetros caracterizam ou especificam a densidade de uma amostra e a carga dos parâmetros nominais, respectivamente. [Leiva et al. \(2008b\)](#) fizeram uma análise baseada na função de risco da BS generalizada (gbs). [Leiva et al. \(2008a\)](#) trouxeram aplicabilidade para a BS generalizada demonstrando que era útil para descrever dados ambientais. [Leiva et al. \(2008c\)](#) buscaram por percorrer sobre desenvolvimento de números aleatórios da BS generalizada. [Barros, Paula e Leiva \(2009\)](#) desenvolveram um pacote em R para a Birnbaum-Saunders generalizada (gbs) que pode ser utilizado para analisar dados a partir de modelos cuja distribuição BS generalizada se adequada.

Também já se encontra na literatura a versão bivariada da BS que foi introduzida por [Kundu, Balakrishnan e Jamalizadeh \(2010\)](#) que mostraram propriedades da BS bivariada, função de densidade conjunta, marginais e condicionais, além de algoritmo para geração de variáveis aleatórias bivariadas. Estimadores de máxima verossimilhança também são apresentados. [Kundu, Balakrishnan e Jamalizadeh \(2013\)](#) apresentaram uma versão multivariada da distribuição BS baseada em uma família de distribuições elípticas

e, recentemente, [Vilca, Balakrishnan e Zeller \(2014b\)](#) estudaram sobre uma extensão da BS bivariada baseada em uma classe de distribuições mistura de escala normal. Em [Vilca, Balakrishnan e Zeller \(2014a\)](#) é apresentada uma outra distribuição relacionada com a BS.

Recentemente, [Santos-Neto *et al.* \(2012\)](#) propuseram uma nova parametrização para a distribuição BS, chamada de RBS, baseada na média de tal forma que o parâmetro de escala é a própria média trazendo facilidades na obtenção dos estimadores dos parâmetros, isso se Y segue uma distribuição Birnbaum-Saunders reparametrizada, denotada por $Y \sim RBS(\mu, \delta)$, μ representa a média e δ o parâmetro de precisão. Além disso em [Santos-Neto *et al.* \(2012\)](#) os autores apresentam um modelo que generaliza o modelo proposto por [Birnbaum e Saunders \(1969a\)](#). Em [Leiva *et al.* \(2014a\)](#) os autores propõem um modelo de regressão baseado na distribuição RBS, em que a variável resposta original não necessita de transformação. Ainda no contexto de distribuições multivariadas [Saulo *et al.* \(2018\)](#) consideraram a reparametrização proposta por [Santos-Neto *et al.* \(2012\)](#) para aplicação em confiabilidade.

Muitos casos reais existe a presença de zeros maior do esperada para modelos estatísticos que são comumente utilizados, como por exemplo, informações referentes a inadimplência do mercado financeiro (de uma instituição financeira específica) em que observamos uma alta ocorrência de 0 na variável resposta. Também observamos estas características em estudos que envolvem doenças, uma vez que se pensarmos em uma específica ocorrência (Resposta = 1) na população é baixa fazendo com que o complementar desta seja elevada (Resposta = 0). Além disso, precisamos levar em conta que nestes casos a resposta nunca poderá assumir valores negativos o que faz com que as abordagens atuais considerando normalidade dos dados não é a mais adequada, salvo os casos em que se utiliza regressão logística para modelagem. Diante desta importância em se trabalhar com uma distribuição de probabilidade que assuma somente valores não negativos e que também acomode, de uma maneira adequada, este conhecimento *a priori* da concentração (inflação) de zeros se faz importante. Na literatura podemos encontrar várias referências que utilizam modelos aditivos (ou de mistura) para se trabalhar em situações na qual existe probabilidade positiva no zero. Neste contexto, uma nova classe de modelo foi estudada em [Tomazella *et al.* \(2019\)](#): a distribuição RBS Zero Ajustada (ZARBS).

A parametrização da distribuição BS apresentada em Santos-Neto *et al.* (2012) nos permite imitar uma propriedade da distribuição gama (versatilidade para acomodar diferentes curvas de risco), a qual foi a primeira distribuição usada em um modelo de fragilidade (LEÃO *et al.*, 2018). Leiva *et al.* (2016) propuseram uma metodologia para logística de estoque que permite modelar dados de demanda com zeros por meio de uma nova mistura de distribuição discreta-contínua, que é construída usando uma massa de probabilidade em zero e componente contínuo relacionado à distribuição RBS. Eles nomearam essa nova classe de modelos como a distribuição RBS ajustada a zero (ZARBS). Com base neste modelo, Tomazella *et al.* (2019), propuseram uma classe geral de modelos de regressão para modelagem de variável aleatória mista. Este novo modelo generaliza pelo menos sete modelos de regressão Birnbaum-Saunders reparametrizados. A especificação proposta para o modelo de regressão ZARBS segue a estrutura tradicionalmente usada em modelos contínuos ajustados a zero.

Neste trabalho, apresentaremos informações importantes para birnbaum-Saunders Reparametrizada (RBS), tais como: gráficos da função de densidade, função de distribuição, os estimadores dos parâmetros pelos métodos dos Momentos (MM) e o Método da Máxima Verossimilhança (MV). Uma outra maneira de caracterizar a distribuição é usar sua função característica que nos permite obter seus momentos. Na literatura sobre a distribuição BS, a função característica é praticamente não estudada. Além disso, estudaremos a distribuição ZARBS sob enfoque clássico e Bayesiano.

1.1 Objetivo

A distribuição Birnbaum-Saunders reparametrizada (RBS) tem mostrado um bom potencial para o desenvolvimento de novas metodologias. Este trabalho tem como principal contribuição desenvolver uma análise Bayesiana baseada no modelo proposto por Tomazella *et al.* (2019). Neste contexto será estudado o pacote BAMLSS-Modelo aditivo Bayesiano para Posição, Escala, Forma desenvolvido em linguagem R (ver (R Core Team, 2021)). A metodologia proposta será comparada com a análise clássica e aplicada a um conjunto de dados reais.

1.2 Organização do Trabalho

A organização deste trabalho se dá na seguinte forma: No Capítulo 2, apresentamos a distribuição Birnbaum-Saunders na sua forma original, seguimos com a forma reparametrizada proposta em Santos-Neto *et al.* (2012) e introduzimos a distribuição Birnbaum-Saunders reparametrizada com zero ajustada (ZARBS). Além disso, apresentamos as ideias básicas de inferência Bayesiana. No Capítulo 3, apresentamos as ferramentas computacionais para a aplicabilidade do modelo de regressão ZARBS. No Capítulo 4, apresentamos o modelo de regressão ZARBS. No Capítulo 5, apresentamos a abordagem Bayesiana para o modelo de regressão ZARBS e análise de diagnóstico para o modelo. Por fim, apresentamos as conclusões e considerações finais no Capítulo 6.

PRELIMINARES

2.1 A Distribuição Birnbaum-Saunders (BS)

Características importantes da distribuição BS serão apresentadas neste capítulo, bem como a ideia inicial de sua formulação. Birnbaum e Saunders ([BIRNBAUM; SAUNDERS, 1969a](#)) fizeram algumas suposições sobre o processo de fadiga, são elas:

- Um material é submetido a uma padrão cíclico de tensão e força;
- A falha do material ocorre devido ao desenvolvimento e ao crescimento de uma fissura dominante dentro do material, ou seja, quando o tamanho da fissura excede certo nível de resistência, denotado por ω ;
- A sequência de tensão imposta ao material é constante de ciclo para ciclo;
- A extensão incremental da fissura X_i resultante da aplicação da i -ésima oscilação de carga é uma variável aleatória com uma distribuição que só depende da fissura atual causada pela tensão neste ciclo;
- A extensão da fissura durante o $(j + 1)$ -ésimo ciclo é

$$Y_{j+1} = X_{j+1} + \dots + X_{j+m} \quad j = 0, 1, \dots \quad (2.1)$$

em que X_{j+i} é a extensão da fissura após a i -ésima oscilação de carga do $(j+1)$ -ésimo ciclo;

- As extensões das fissuras em diferentes ciclos são diferentes;
- A extensão total da fissura, Y_j , $j = 1, 2, \dots$, devido ao j -ésimo ciclo é uma variável aleatória independente e identicamente distribuída com média μ e variância σ^2

2.1.1 Funções Densidade e Distribuição de Probabilidade

Dadas as condições supracitadas e mais alguns detalhes, [Birnbaum e Saunders \(1969a\)](#) mostraram que para uma variável aleatória T independente e identicamente distribuída, a função de densidade de probabilidade é dada da seguinte forma:

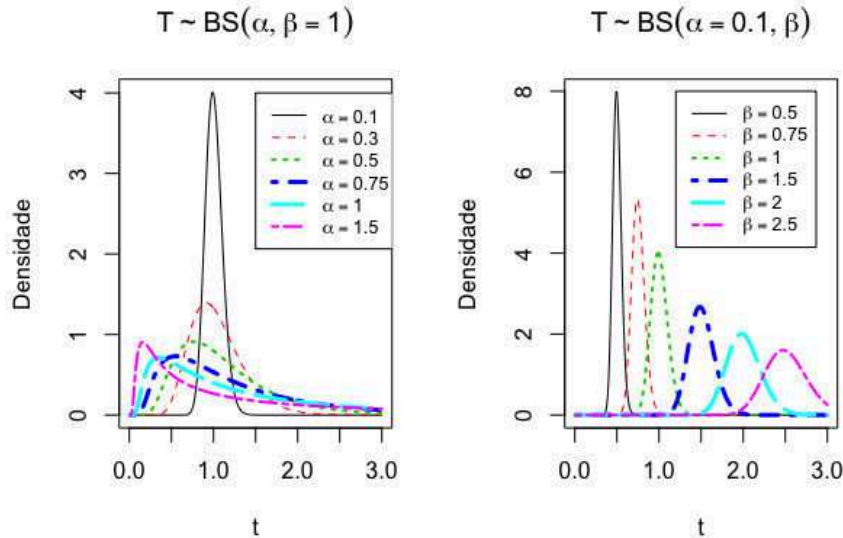
$$f(t; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right) \frac{[t + \beta]}{2\alpha\sqrt{\beta t^3}}, \quad t > 0. \quad (2.2)$$

Se uma variável aleatória absolutamente contínua T possui densidade como a apresentada em (2.2), então dizemos que T tem distribuição Birnbaum-Saunders, com parâmetro de forma $\alpha > 0$ e parâmetro de escala $\beta > 0$, em que denotamos por $T \sim BS(\alpha, \beta)$. Além disso, esta distribuição tem algumas propriedades importantes. Dentre elas, temos:

- Se $T \sim BS(\alpha, \beta)$, então $cT \sim BS(\alpha, c\beta)$ com $c > 0$ e;
- Se $T \sim BS(\alpha, \beta)$, então $\frac{1}{T} \sim BS\left(\alpha, \frac{1}{\beta}\right)$.

A Figura 2.1.1 nos mostra o comportamento da função densidade de probabilidade para diferentes valores dos parâmetros da distribuição BS. O gráfico da esquerda aponta uma alteração na forma da distribuição de maneira que a medida que o α aumenta a distribuição se torna mais assimétrica. Já o gráfico da direita da mesma Figura 2.1.1 mostra que a medida que β aumenta a distribuição é deslocada e seu ponto de maior massa é exatamente o valor de β .

Figura 2.1.1 – Função de densidade da BS para diferentes parâmetros.



Fonte: Elaborada pelo autor.

Também é possível caracterizar a distribuição Birnbaum-Saunders através da sua função de distribuição acumulada (f.d.a). Portanto, se uma variável aleatória T possui distribuição Birnbaum-Saunders, então sua f.d.a pode ser escrita como:

$$F_T(t; \alpha, \beta) = P(T \leq t) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right], \quad t > 0, \quad (2.3)$$

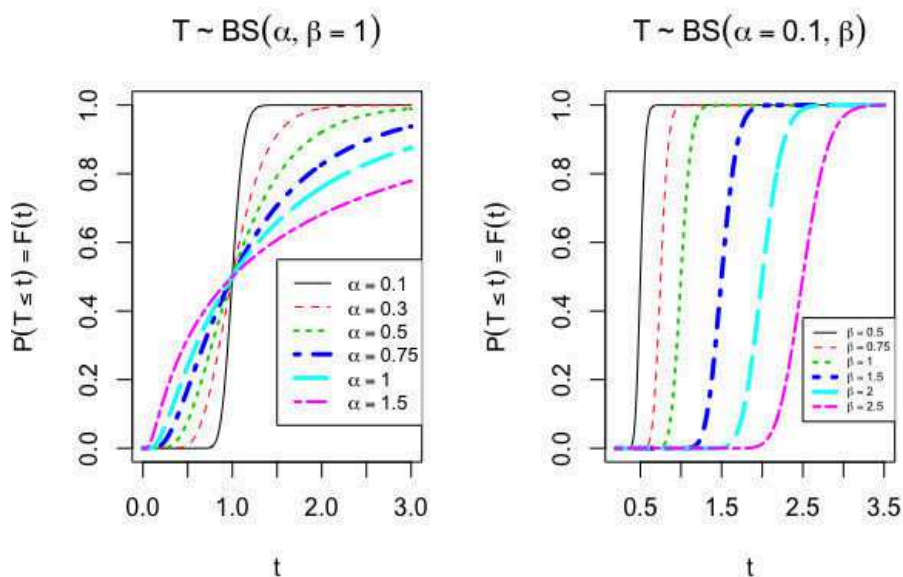
em que $\Phi(\cdot)$ é a f.d.a de uma distribuição normal padrão. Além disso, pode-se demonstrar que o parâmetro β é a mediana da distribuição, pois $F(\beta) = 0.5$. Na Figura 2.1.2, apresentamos o comportamento da função de distribuição acumulada para diferentes combinações dos parâmetros α e β .

A média e a variância do modelo BS são dadas, respectivamente, por

$$E[T] = \beta \left[1 + \frac{\alpha^2}{2} \right] \quad \text{e} \quad \text{Var}[T] = [\beta\alpha]^2 \left[1 + \frac{5\alpha^2}{4} \right].$$

Note, também, que se $\alpha \rightarrow 0$, então a Birnbaum-Saunders converge para uma distribuição degenerada em β .

Figura 2.1.2 – Função de Distribuição para diferentes parâmetros.



Fonte: Elaborada pelo autor.

As medidas de assimetria e a curtose (denotados por μ_3 e μ_4 , respectivamente) podem ser escritas das seguintes maneiras:

$$\mu_3 = \frac{16\alpha^2(11\alpha^2 + 6)}{(5\alpha^2 + 4)^3} \quad \text{e} \quad \mu_4 = 3 + \frac{6\alpha^2(93\alpha^2 + 41)}{9 + \alpha^2},$$

sendo possível notar que o parâmetro α controla o achatamento e a assimetria da distribuição. Além das medidas apresentadas acima, uma medida de dispersão relativa e que pode ser de interesse prático é o coeficiente de variação (CV). Para o modelo Birnbaum-Saunders, $BS(\alpha, \beta)$, temos que essa medida é dada por

$$CV = \frac{\sqrt{5\alpha^4 + 4\alpha^2}}{\alpha^2 + 2}.$$

2.2 Distribuição Birnbaum-Saunders Reparametrizada

Santos-Neto *et al.* (2012) propuseram uma reparametrização da distribuição Birnbaum-Saunders. Nesta versão, a Birnbaum-Saunders é indexada pela sua média e por um parâmetro de precisão. Com isso, foi possível obter alguns ganhos matemáticos como a obtenção do estimadores de momentos e componente do desvio, ver Santos-Neto *et al.* (2014) e Leiva *et al.* (2014b).

Esta reparametrização da distribuição BS foi obtida considerando que $\alpha = \sqrt{\frac{2}{\delta}}$ e $\beta = \frac{\delta\mu}{\delta+1}$, e a função densidade de probabilidade pode ser escrita da seguinte maneira

$$f(y; \delta, \mu) = \frac{e^{\frac{\delta}{2}} \sqrt{\delta+1}}{4\sqrt{\pi\mu}y^{\frac{3}{2}}} \left[y + \frac{\delta\mu}{\delta+1} \right] \exp \left(-\frac{\delta}{4} \left[\frac{y(\delta+1)}{\delta\mu} + \frac{\delta\mu}{y(\delta+1)} \right] \right), \quad y > 0, \quad (2.4)$$

em que $\delta > 0$ e $\mu > 0$ são, respectivamente, os parâmetros de precisão e a média. A versão reparametrizada da distribuição Birnbaum-Saunders conserva algumas das principais propriedades.

De Santos-Neto *et al.* (2014) temos que a média e a variância são respectivamente dadas por

$$E[Y] = \mu \quad \text{e} \quad V[Y] = \frac{2\delta+5}{[\delta+1]^2}.$$

A função de distribuição acumulada e a função quantílica podem ser expressas, respectivamente, como a seguir:

$$F(y; \mu, \delta) = \Phi \left(\sqrt{\frac{\delta}{2}} \left[\sqrt{\frac{(\delta+1)y}{\mu\delta}} - \sqrt{\frac{\mu\delta}{(\delta+1)y}} \right] \right) \quad y > 0, \quad (2.5)$$

e

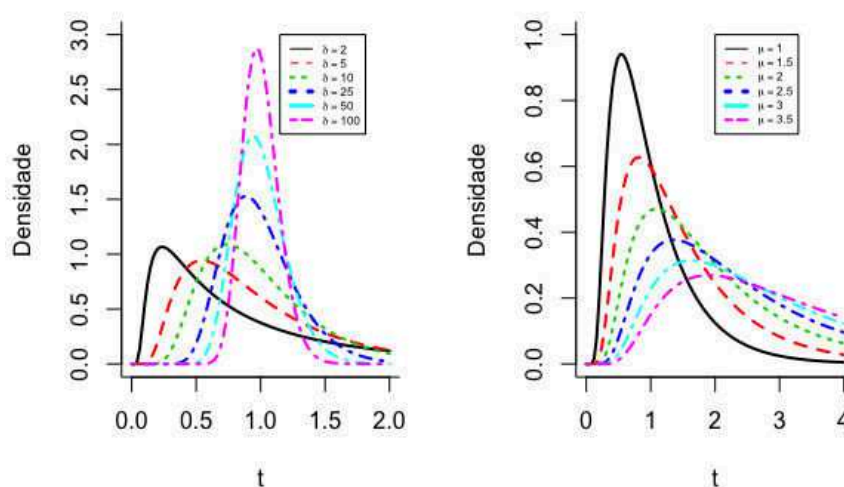
$$y(q; \mu, \delta) = F^{-1}(q) = \frac{\delta\mu}{\delta+1} \left[\frac{z(q)}{\sqrt{2\delta}} + \sqrt{\left[\frac{z(q)}{\sqrt{2\delta}} \right]^2 + 1} \right]^2, \quad 0 < q < 1, \quad (2.6)$$

em que $z(q)$ representa o q -ésimo quantil da normal padrão e F^{-1} é o inverso da FDA de Y . Assim, temos que função de risco de Y é dada por:

$$h(y; \mu, \delta) = \frac{\exp\left(\frac{\delta}{2}\right) \sqrt{\delta+1}}{4\sqrt{\pi\mu y^3}} \left[y + \frac{\delta\mu}{\delta+1} \right] \frac{\exp\left(\frac{\delta}{4} \left[\frac{(\delta+1)y}{\delta\mu} + \frac{\delta\mu}{(\delta+1)y} \right]\right)}{\Phi\left(-\sqrt{\frac{\delta}{2}} \left[\sqrt{\frac{(\delta+1)y}{\mu\delta}} - \sqrt{\frac{\delta\mu}{(\delta+1)y}} \right]\right)}. \quad (2.7)$$

A Figura 2.2.1 nos mostra o comportamento da função densidade de probabilidade para diferentes valores dos parâmetros da distribuição RBS. Dos quais podemos extrair as seguintes informações: a) a medida que os valores de δ aumentam, a distribuição se torna mais simétrica; b) A medida que os valores de μ aumentam, percebemos uma maior dispersão dos dados.

Figura 2.2.1 – Função Densidade de Probabilidade da variável aleatória RBS para diferentes valores dos parâmetros



Fonte: Elaborada pelo autor.

2.3 Inferência Bayesiana

A abordagem Bayesiana é um tipo de inferência estatística que descreve as incertezas sobre quantidades desconhecidas de forma probabilística. Os graus de incerteza são representados através de modelos probabilísticos para os parâmetros. Neste contexto, é natural que diferentes pesquisadores possam ter diferentes graus de incerteza a respeito dos parâmetros. Sendo assim, não existe nenhuma distinção entre quantidades observáveis e os parâmetros de um modelo estatístico, todos são considerados quantidades aleatórias. Incertezas são modificadas periodicamente após observações de novos dados ou resultados.

A operação que calibra a medida das incertezas é conhecida como operação Bayesiana e é baseada na fórmula de Bayes. A fórmula de Bayes é muitas vezes denominada Teorema de Bayes. Este teorema nos trás uma informação importante que é a de que o espaço amostral pode ser atualizado (reduzido) ao sabermos de uma nova informação que antes não conhecíamos. Bayes (1958) iniciou a abordagem dos dados levando em consideração a informação dos próprios dados e atualização por conhecimentos *a priori*.

Assim coleta-se uma amostra provenientes da população de interesse, X , com a qual o parâmetro de interesse θ pode ser obtido. A relação entre estas informações sobre os dados e o parâmetro desconhecido é dada pelo teorema de Bayes, descrito em 2.8.

$$p(\theta | x) = \frac{p(x, \theta)}{p(x)} = \frac{p(x | \theta) p(\theta)}{p(x)} = \frac{p(x | \theta) p(\theta)}{\int p(\theta, x) d\theta}. \quad (2.8)$$

Vale ressaltar que $\frac{1}{p(x)}$ não depende de θ e, portanto, é considerada como uma constante normalizadora de $p(\theta | x)$.

Para recuperar a constante normalizadora basta seguir:

$$p(x) = \int p(x, \theta) d\theta = \int p(x | \theta) p(\theta) d\theta = E_{\theta} [p(X | \theta)]. \quad (2.9)$$

Para um valor fixo de x a função $L(x | \theta) = p(x | \theta)$ fornece a verossimilhança de cada um dos possíveis valores de θ , enquanto que $p(\theta)$ é chamada de distribuição *a priori* de θ . Com estas duas informações combinadas que conseguimos chegar na função de interesse que é conhecida como distribuição *a posteriori* de θ . A forma usual da

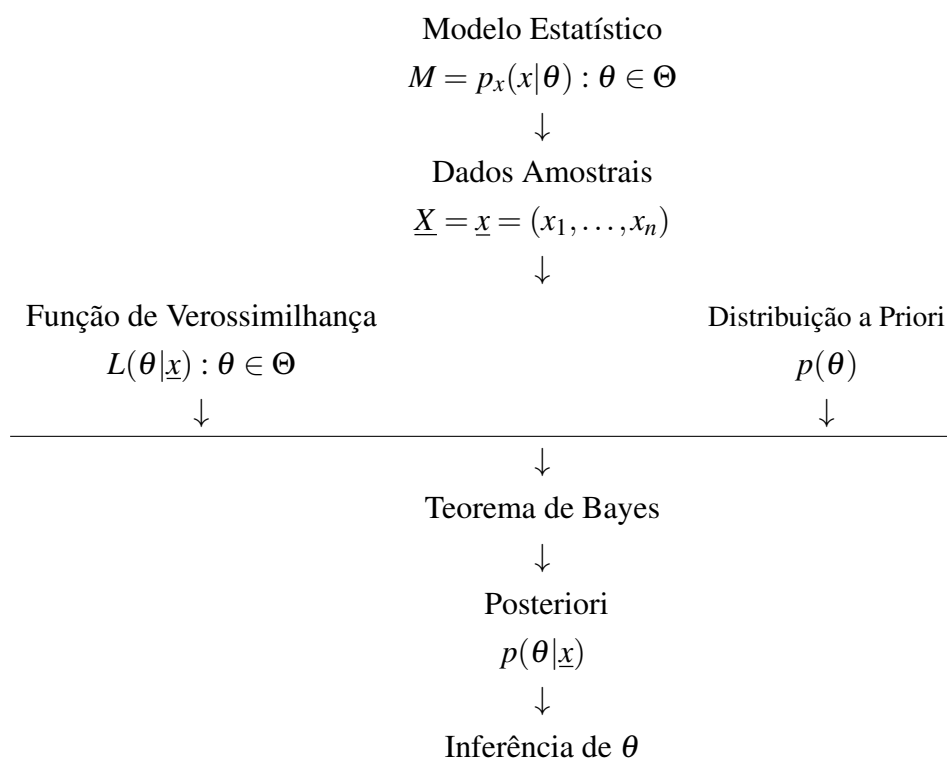
posteriori definida por Bayes é dada por:

$$p(\theta | x) \propto L(\theta; x) p(\theta). \quad (2.10)$$

Esta forma de escrever é bastante prática para o caso de estimação de parâmetros, na qual a constante serve apenas como normalizadora. Em outras situações obter esta constante é importante para solução do problema.

A distribuição *a posteriori* $p(\theta | x)$ expressa todo o conhecimento que temos a respeito do parâmetro, depois de observarmos os dados (que dependem do parâmetro). Isto significa que a probabilidade $p(\theta)$ será revista à luz dos dados x , originando $p(\theta | x)$. Ou seja, os dados tendem a corrigir a informação inicial, embora essa idéia não significa necessariamente que a informação estivesse equivocada: ela poderia estar incompleta, e é o indivíduo que aos poucos vai incorporando mais e mais informação, a ponto de ir mudando de opinião para tirar suas conclusões. Desta forma, conhecendo a distribuição do parâmetro, podemos examinar qualquer aspecto de θ (média, variância, quantis, probabilidade de assumir determinados valores, etc.).

O diagrama a seguir mostra o mecanismo de como obter a distribuição *a posteriori*.



A escolha das estimativas Bayesianas dependem da forma da *posteriori* e (do objetivo) e as estimativas mais utilizadas são a moda *a posteriori*, a média *a posteriori* e a mediana *a posteriori*.

Moda a posteriori

$$\theta_1^* \text{ é tal que } p(\theta^*|\mathbf{x}) = \max_{\theta \in \Theta} p(\theta|\mathbf{x}) = \max_{\theta \in \Theta} \{p(\theta)L(\theta|\mathbf{x})\}$$

Média a posteriori

$$\theta_2^* = E\{\theta_i|\mathbf{x}\} = \int_{\Theta} \theta_i p(\theta|\mathbf{x}) d\theta, \quad i = 1, \dots, k.$$

Mediana a posteriori (vetor)

$$P[\theta_i \geq \theta_3^*|\mathbf{x}] \geq \frac{1}{2} \quad \text{e} \quad P[\theta_i \leq \theta_3^*|\mathbf{x}] \geq \frac{1}{2}, \quad i = 1, \dots, k.$$

Também podemos obter via estimação por região de credibilidade, construir testes de Hipóteses calculando as respectivas probabilidades a posteriori (Fator de Bayes).

A distribuição de uma nova observação, partindo da ideia de que $p(x)p(\theta|x) = p(x|\theta)p(\theta)$ é dado por:

$$p(x) = \frac{p(x|\theta)p(\theta)}{p(\theta|x)} = \frac{a(x)k(\alpha, \beta)}{k(\alpha + u(x), \beta + 1)}. \quad (2.11)$$

A implementação de técnicas Bayesianas é adequada para modelos complexos de difícil tratamento analítico, como por exemplo, o grande número de parâmetros a serem estimados, a dificuldade na obtenção das densidades marginais de forma analítica, ou também quando as distribuições *a priori* e *a posteriori* são conjugadas. Nesses casos, os métodos analíticos de aproximação não são indicados, e métodos de aproximação numérica são necessários para estimação dos parâmetros de interesse.

2.3.1 Distribuições a Priori

Em inferência Bayesiana, a distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre θ antes da realização do experimento

A distribuição *a priori* pode ser considerada de diferentes formas:

- pode ser elaborada a partir de opiniões de especialistas;
- pode ser baseada em dados existentes, por exemplo, uma meta-análise;
- sob vários procedimentos objetivos;
- para obtermos conjugação, que é útil mas não necessária;
- pode ser suposta desconhecida e estimada usando simulações similares repetidas.

Prioris Conjugadas

Levando em consideração a informação que se tem sobre θ , pode-se definir uma família paramétrica de distribuições. A priori passa a ser representada por uma forma funcional na qual os parâmetros são especificados de acordo com este conhecimento. Para diferenciar o parâmetro de interesse e os diversos parâmetros da distribuição, estes últimos são chamados de *hiperparâmetros*. Neste caso, a ideia é a de que tanto a priori quanto a posteriori pertencem a mesma família de distribuições e, por conta disso, a atualização que se tem sobre θ envolve apenas uma atualização nos hiperparâmetros. De maneira matemática temos:

Definição 1. Se $F = p(x | \theta)$, $\theta \in \Theta$ é uma classe de distribuições amostrais. Então uma classe de distribuições P é conjugada a F se

$$\forall p(x | \theta) \in F \text{ e } p(\theta) \in P \Rightarrow p(\theta | x) \in P. \quad (2.12)$$

Principais famílias conjugadas são:

- Binomial: a família Beta é conjugada à Binomial (ou Bernoulli);
- Poisson: a família gama é conjugada à Poisson.
- Normal com variância conhecida: a família normal é conjugada à normal;

Exemplo:

Seja $X \sim \text{bin}(n, \theta)$. Assuma que $\theta \sim \text{beta}(a, b)$: *priori* para θ , com a e b conhecidos:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \dots, n,$$

e

$$\pi(\theta) = \frac{1}{B(a, b)} \theta^{a-1} (1-\theta)^{b-1}, \quad 0 \leq \theta \leq 1,$$

A *posteriori* para θ , $\pi(\theta|x)$ é dada por,

$$\pi(\theta|x) = \frac{\theta^{x+a-1} (1-\theta)^{n+b-x-1}}{\int_0^1 \theta^{x+a-1} (1-\theta)^{n+b-x-1} d\theta}.$$

$$\pi(\theta|x) \propto \theta^{x+a-1} (1-\theta)^{n+b-x-1},$$

em que $k = \int_0^1 \theta^{x+a-1} (1-\theta)^{n+b-x-1} d\theta$.

Como

$$k^{-1} = \int_0^1 \theta^{x+a-1} (1-\theta)^{n+b-x-1} d\theta = \frac{\Gamma(x+a)\Gamma(n+b-x)}{\Gamma(n+a+b)},$$

temos que,

$$\pi(\theta|x) = \frac{1}{B(x+a, n+b-x)} \theta^{x+a-1} (1-\theta)^{n+b-x-1},$$

para $0 \leq \theta \leq 1$. Note que a distribuição *a posteriori* para θ é dada pela distribuição beta.

Se existe k tal que $k^{-1} = \int L(\theta|x)\pi(\theta) < \infty$ e todo $\pi \in \Psi$ é definido através de $\pi(\theta|x) = k^{-1}L(\theta|x)\pi(\theta)$. Então, Ψ é dita ser da família conjugada natural ao modelo amostral gerador de $L(\theta|x)$.

2.3.1.1 Conjugação na Família Exponencial

Fazer parte da família exponencial é algo que facilita muito a resolução de problemas envolvendo uma distribuição de probabilidade, por conta da fácil obtenção de estatísticas e suas propriedades. Felizmente, muitas das distribuições mais usadas na estatística fazem parte desta família de distribuições, tanto distribuições contínuas quanto discretas. Por fazer parte desta família, a distribuição possui uma estatística suficiente com dimensão fixa.

Definição 2. A família de distribuição com função de densidade de probabilidade $p(x | \theta)$ pertence a família exponencial de um parâmetro se podemos escrever:

$$p(x | \theta) = a(x) \exp[u(x) \phi(\theta) + b(\theta)]. \quad (2.13)$$

Vale ressaltar que pelo critério de fatoração de Neyman $U(x)$ é uma estatística suficiente para θ . Sendo assim, a classe conjugada é identificada como:

$$p(\theta) = k(\alpha, \beta) \exp[\alpha \phi(\theta) + \beta b(\theta)]. \quad (2.14)$$

Fazendo aplicação do Teorema de Bayes chegamos a:

$$p(\theta | x) = k(\alpha + u(x), \beta + 1) \exp[[\alpha + u(x)] \phi(\theta) + [\beta + 1]]. \quad (2.15)$$

Exemplo Seja $X \sim \text{bin}(n, \theta)$. Assuma que $\theta \sim \text{beta}(r, s)$:

$$p(x|\theta) = \binom{n}{x} \exp \left\{ x \log \left(\frac{\theta}{1-\theta} \right) + n \log(1-\theta) \right\}$$

$$p(x|\theta) = a(x) \{T(x)c(\theta) + d(\theta)\}.$$

A família conjugada é

$$\begin{aligned} \pi(\theta) &\propto \theta^{r-1} (1-\theta)^{s-1} \\ &\propto \exp \left\{ (r-1) \log \left(\frac{\theta}{1-\theta} \right) + \left(\frac{s+r-2}{n} \right) n \log(1-\theta) \right\} \\ &\propto \exp \{ \alpha c(\theta) + \beta d(\theta) \}. \end{aligned}$$

A posteriori também é Beta com parâmetros $r+x$ e $s+n-x$

$$\begin{aligned}\pi(\theta|x) &\propto \exp\left\{(r+x-1)c(\theta) + \left(\frac{s+r-2+n}{n}\right)nd(\theta)\right\} \\ &\propto \theta^{r+x-1}(1-\theta)^{s+n-x-1}.\end{aligned}$$

2.3.1.2 Priori não-informativa

Na prática, podemos não ter informação a priori sobre um parâmetro de interesse. Isto refere-se a especificação de distribuições a priori quando se espera que a informação dos dados seja dominante, no sentido de que a nossa informação a priori é vaga. Os conceitos de conhecimento “vago”, “não Informação”, ou “ignorância a priori” claramente não são únicos e o problema de caracterizar prioris com tais características pode se tornar bastante complexo.

A principal ideia quando se trata de *prioris* não informativas é levar em consideração mais informações com respeito aos dados do que com respeito à *priori* e, desta maneira, considerar que os valores de θ são equiprováveis, ou seja, determinar uma distribuição uniforme. Matematicamente falando, seria considerar $p(\theta) \propto k$ para θ variando em um subconjunto da reta. Desta maneira, nenhum valor de θ tem preferência (BAYES, 1763). Vale lembrar que é possível na abordagem Bayesiana que esta priori possa ser definida pelo próprio pesquisador que detem conhecimento a respeito do comportamento dos seus dados.

Algumas formas de prioris não-informativas são:

- Uniforme
- Box-Tiao
- Jeffreys
- Máxima Entropia

Priori Uniforme

A primeira idéia de não informação a priori que se pode ter é pensar em todos os possíveis valores de θ como igualmente prováveis, i.e., com uma distribuição à priori uniforme. Neste caso,

$$\pi(\theta) \propto k,$$

para θ variando em um subconjunto da reta, o que significa que nenhum valor particular tem preferência.

Se o intervalo de variação de θ for ilimitado então a distribuição θ é imprópria, i.e.

$$\int \pi(\theta) d\theta = \infty.$$

Se $\phi = g(\theta)$ é uma reparametrização não linear monótona de θ , então $\pi(\phi)$ é não uniforme já que pelo teorema de transformação de variáveis

$$\pi(\phi) = \pi(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|.$$

Na prática, como estaremos interessados na distribuição a posteriori não daremos muita importância à impropriedade da distribuição à priori. Por outro lado, devemos sempre nos certificar de que a posterior é própria antes de fazer qualquer inferência.

Método de Box-Tiao

A idéia básica foi procurar uma reparametrização $\phi = \phi(\theta)$ do modelo $\{p(x|\theta), \theta \in \Theta\}$ para o qual a respectiva verossimilhança fosse apenas transladada pelos dados, isto é.

$$L(\theta|x) \approx g[\phi(\theta) - m(x)].$$

onde g é uma função cuja forma é independente de x e $m(x)$ a função que descreve a translação de L com variação de x (o parâmetro de locação de L).

Uma vez detectada uma transformação ϕ deste tipo, consideramos para ela uma distribuição que assegure que a respectiva distribuição a posteriori seja essencialmente a verossimilhança normalizada, em termos aproximados temos

$$\pi(\phi|x) \approx \frac{L(\phi|x)}{\int L(\phi|x)}.$$

A distribuição não-informativa de Box-Tiao para o parâmetro original θ é aproximadamente proporcional ao jacobiano da transformação que se representa por:

$$\pi(\theta) \propto \left| \frac{d\phi}{d\theta} \right|.$$

Priori de Jeffreys

O caráter não informativo da distribuição a priori gerado pela regra de Jeffreys vem do fato de ser obtida a partir do modelo gerador de dados.

O uso da priori de Jeffreys justifica-se pela propriedade de invariância e baseia-se na utilização da informação Fisher de $\theta \in R$

$$I(\theta) = E \left[-\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} \right]. \quad (2.16)$$

E se θ é um vetor de parâmetros então $\pi(\theta) \propto |I(\theta)|^{1/2}$ e (2.16) é dado por,

$$I(\theta) = E \left[-\frac{\partial^2 \log p(x|\theta)}{\partial \theta \partial \theta'} \right].$$

Considerando o caso de um parâmetro para qualquer transformação nós temos

$$I(\phi) = I(\theta(\phi)) \left(\frac{d\theta}{d\phi} \right)^2. \quad (2.17)$$

Assim, a priori Jeffreys tem a propriedade de invariância.

$$\pi(\phi) = \pi(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right|.$$

No caso multiparamétrico, temos o mesmo argumento em torno da matriz de informação de Fisher $I(\theta)$ com elementos

$$I_{ij}(\theta) = E \left[\frac{\partial \ln p(x|\theta)}{\partial \theta_i} \frac{\partial \ln p(x|\theta)}{\partial \theta_j} \middle| \theta \right].$$

Assim a priori de Jeffreys é dada por

$$\pi(\theta) \propto |\det I(\theta)|^{1/2}.$$

Para qualquer transformação de um -a-um, temos

$$I(\varphi) = \left(\frac{\partial \theta}{\partial \varphi} \right)' I(\theta(\varphi)) \frac{\partial \theta}{\partial \varphi}. \quad (2.18)$$

A priori de Jeffreys aplicada ao caso multiparamétrico normalmente tem implicações que são muitas vezes insatisfatórias ultrapassadas pela imposição a priori de independência entre os parâmetros e uso da priori de Jeffreys separada para especificações das distribuições marginais.

Método de Máxima Entropia

Jaynes(1968) sugeriu a ideia de usar o conceito de entropia, o qual é utilizado em física como uma medida da quantidade de desordem e imprevisibilidade de um sistema físico.

Definição (Caso Discreto): Seja θ um parâmetro discreto com função massa de probabilidade $\pi(\theta)$ e suporte Θ . Definimos entropia de $\pi(\theta)$ como sendo o valor esperado de $-\ln \pi(\theta)$ ou seja

$$E(\pi(\theta)) = - \sum_{\theta \in \Theta} \ln(\pi(\theta)) \pi(\theta).$$

- Ideia: Usar $E(\pi(\theta))$ para encontrar uma priori não-informativa para θ

Definição (Caso Contínuo): Definimos entropia de uma distribuição $\pi(\theta)$, $\theta \in R$, como

$$E(\pi(\theta)) = - \int (\ln(\pi(\theta)))\pi(\theta)d\theta.$$

Nota: Baseado em cálculo das variações, se θ for definido num intervalo limitado da reta R , a distribuição $\pi(\theta)$ que maximiza $E(\pi(\theta))$ sujeito a restrição $\int \pi(\theta)d\theta = 1$, é a distribuição uniforme.

Observar que a função $y = \pi(\theta)$ que maximiza

$$- \int y \ln y d\theta + \lambda_0 \left(\int_{\Theta} y d\theta - 1 \right),$$

é solução de $\frac{\partial F}{\partial y} = 0$ com $F(y, \theta) = -y \ln y + \lambda_0 y$, resultando que $\pi(\theta)$ tem de ser constante.

Nota: Jaynes(1968) redefine entropia de uma distribuição $\pi(\theta)$ no caso contínuo como

$$E(\pi(\theta)) = - \int \pi(\theta) \ln \frac{\pi(\theta)}{\pi_0(\theta)}.$$

Em que $\pi_0(\theta)$ é uma distribuição a priori de referência não-informativa

Se admitirmos, como no caso discreto, a existência de informação inicial representada por $E(g_j(\theta)) = \mu_j$, $j = 1, \dots, m$, para m funções g_j bem definidas, podemos provar que a função de densidade de probabilidade $y = \pi(\theta)$ que maximiza a entropia sujeita as restrições impostas é solução de $\frac{\partial F}{\partial y} = 0$, resultando em

$$\pi(\theta) \propto \pi_0(\theta) \exp \left\{ \sum_{i=1}^m \lambda_j g_j(\theta) \right\}.$$

Além dessas prioris não-informativas apresentadas acima, podemos citar a Priori de referência. Considerar priori de referência na análise Bayesiana é um método de produzir afirmações de inferência Bayesiana no qual somente depende do modelo assumido e dos dados observados. A idéia básica é:

- Identificar a forma matemática de uma priori não-informativa.

- Uso de teoria de informação para medir a quantidade de informação sobre o parâmetro de interesse.
- Define a falta de informação sobre a quantidade de interesse
- Define a priori de referência como aquela que maximiza a falta de informação sobre a quantidade interesse.

No entanto, muitos mecanismos usam diferentes configurações do modelo e formatos de resultados, dificultando a comparação de propriedades de diferentes algoritmos ou para selecionar a distribuição apropriada e variáveis, etc. As razões são múltiplas: o uso de diferentes linguagens de especificação de modelos BUGS (LUNN *et al.*, 2009) ou R (R Core Team, 2021); diferentes softwares estatísticos autônomos como BayesX (UMLAUF; KNEIB; KLEIN, 2019), etc. O pacote BAMLSS (UMLAUF; KLEIN; ZEILEIS, 2018a) apresenta um conceito unificado “Lego Tool-box” para modelos de regressão complexos. Os algoritmos de estimação iterativos, por exemplo, para a moda a posteriori ou a estimação da média baseada em simulação MCMC, apresentam uma estrutura muito similar de tal forma que o processo de construção do modelo se torne relativamente simples, já que a arquitetura do modelo é apenas uma combinação de blocos simples. Devido a grande semelhança ao GAMLSS (RIGBY; STASINOPOULOS, 2005), o pacote é chamado de BAMLSS (UMLAUF; KLEIN; ZEILEIS, 2018a). No entanto, também engloba muito mais termos de modelos generalizados além de combinações lineares em uma matriz de planejamento com coeficientes de regressão. O pacote por ser explorado de três maneiras: primeiro, para desenvolver rapidamente novos modelos e algoritmos. Segundo, para comparar algoritmos existentes e amostradores. Terceiro, para facilmente integrar implementações existentes. A prova do conceito é dada no correspondente pacote do R BAMLSS (UMLAUF; KLEIN; ZEILEIS, 2018a).

2.3.2 Critérios de seleção de modelos

No contexto bayesiano existem diferentes critérios que podem ser adotados para selecionar o melhor modelo. São exemplos desses critérios: a medida DIC (Deviation Information Criterion) que foi proposto por Spiegelhalter *et al.* (2002), o EAIC (Expected Akaike Information Criterion) que por sua vez foi proposto por Brooks *et al.* (2002) e o

EBIC (Expected Bayesian Information Criterion) proposto por [Carlin e Louis \(2000\)](#). Esses critérios são baseados no desvio médio a posteriori, $\mathbb{E}(D(\theta))$, dados por:

$$DIC = \bar{D} + P_D = 2\bar{D} - \hat{D} \quad (2.19)$$

$$EAIC = \bar{D} + 2b \quad (2.20)$$

$$EBIC = \bar{D} + b \log(n), \quad (2.21)$$

na qual $\bar{D} = \frac{1}{Q} \sum 1 = q^Q D(\theta^{(q)})$ no qual o índice q indica a q -ésima realização de um total de Q realizações e $D(\theta) = -2 \sum_1^n \log(f(y_i | \theta))$ na qual $f(\cdot)$ é a função densidade de probabilidade da distribuição em questão, no nosso caso a ZARBS. Nas equações acima b é o número de parâmetros no modelo e P_D é o número efetivo de parâmetros, definido como: $\mathbb{E}[D(\theta) - D[\mathbb{E}(\theta)]]$, na qual $D[\mathbb{E}(\theta)]$ é o desvio médio a posteriori que pode ser estimado como: $\hat{D} = D\left(\frac{1}{Q} \sum_{q=1}^Q D(\theta^{(q)})\right)$. Em posse dessas medidas, para selecionar o modelo a partir delas basta observar aquele que possui os menores valores para elas.

Outro critério vastamente utilizado para seleção de modelos é a Densidade Preditiva Ordinária (CPO). Em [Gelfand, Dey e Chang \(1992\)](#) e [Gelfand e Dey \(1994\)](#) encontramos discussões detalhadas sobre a estatística CPO e suas aplicações em seleção de modelos.

Seja D o conjunto de dados completo e $D_{(-i)}$ o conjunto de dados com a i -ésima observação deletada. Assim, a densidade a posteriori de θ dado $D_{(-i)}$ para $\pi(\theta | D_{(-i)})$ para $i = 1 \dots n$. Desta maneira, a i -ésima observação do CPO_i pode ser reescrita como:

$$CPO_i = \int_{\Theta} f(y_i | \theta) \pi(\theta | D_{(-i)}) \partial \theta = \left\{ \int_{\Theta} \frac{\pi(\theta | D_{(-i)})}{f(y_i | \theta)} \partial \theta \right\}^{-1}, i = 1, \dots, n, \quad (2.22)$$

na qual $f(y_i | \theta)$ é a função densidade de probabilidade. Altos valores para o CPO_i indica o melhor modelo. Uma estimativa para o CPO_i pode ser obtida usando uma amostra do MCMC para a distribuição a posteriori $\pi(\theta | D)$. Seja $\theta_1, \dots, \theta_Q$ uma amostra de Q de $\pi(\theta | D)$ depois do *burn-in*. Uma aproximação MCMC para o CPO_i ([CHEN; SHAO;](#)

IBRAHIM, 2012) é dada por:

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^n \frac{1}{f(y_i | \theta_q)} \right\}^{-1}. \quad (2.23)$$

Uma estatística resumo do CPO_i é $B = \sum_{i=1}^n \log(CPO_i)$ em que quanto maior o valor de B melhor será o ajuste do modelo.

2.3.3 Conclusão

Neste capítulo apresentamos a distribuição BS, passando pela função densidade de probabilidade, função distribuição, seus estimadores para a média e variância, além de visualizações gráficas com diferentes parâmetros de forma e escala. Seguimos então para a apresentação da RBS (Birnbaum-Saunders reparametrizada), suas propriedades e os estimadores dos parâmetros de média e variância, além das visualizações gráficas que facilitam o entendimento do impacto na alteração dos parâmetros. Por fim, apresentamos uma revisão da abordagem Bayesiana com as diversas maneiras de se definir uma priori, além dos passos até se obter a posteriori (quando combinamos os dados observados junto a priori para determinar a posteriori).

No próximo capítulo detalharemos o procedimento para o uso do pacote BAMLSS. Passaremos pela ideia inicial com modelos lineares generalizados até chegar ao funcionamento do pacote BAMLSS (similar ao GAMLSS) utilizado para modelar dados com a distribuição aqui estudada (Birnbaum-Saunders). Passamos também pelo algoritmo Lego-Bricks desse pacote e todo o seu entendimento até um exemplo explicitando o funcionamento do pacote.

FERRAMENTA COMPUTACIONAL

A abordagem inteiramente Bayesiana usando técnicas de simulação da cadeia de Markov de Monte Carlo (MCMC) é particularmente interessante já que o padrão inferencial produz intervalos válidos e credíveis para os estimadores em situações na qual o intervalo de confiança para os correspondentes estimadores de máxima verossimilhança baseados em propriedades assintóticas falham. Além disso, extensões como seleção de variáveis, prioris não padrões para hiperparâmetros, ou modelos multinível são facilmente inclusos. Por conta disso, e do avanço da capacidade computacional na última década, o número de técnicas de estimação para problemas de inferência, tanto para Bayesianos como para frequentistas, tem recebido cada vez mais atenção.

A maioria das técnicas usam diferentes configurações de modelos e formatos de saída, o que dificulta para os praticantes, como por exemplo comparar prioridades de algoritmos diferentes ou selecionar a distribuição, variáveis adequadas, etc. Os motivos são diversos: o uso de uma linguagem de modelo diferente como *BUGS* (LUNN *et al.*, 2009) ou *R* (R Core Team, 2021); pacotes de softwares estatísticos independentes como BayesX (UMLAUF; KNEIB; KLEIN, 2019), *JAGS* (PLUMMER, 2003), *Stan* (CARPENTER *et al.*, 2017) ou WinBugs (LUNN *et al.*, 2000); ou mesmo diferenças dentro do mesmo ambiente, por exemplo os pacotes *R mgcv* (WOOD, 2003), *GAMLSS* (RIGBY; STASINOPOULOS, 2005), *VGAM* (YEE, 2020) implementam toda a infraestrutura dos termos do modelo em sua própria maneira. Isso é particularmente problemático se

todos os pacotes estão carregados no ambiente global do R, porque algumas funções que supostamente deveriam cumprir o mesmo propósito têm interpretações diferentes.

Neste capítulo será descrito o método baseado no pacote *BAMLSS* (*Bayesian Additive Models for Location, Scale and Shape*), proposto por [Umlauf, Klein e Zeileis \(2018a\)](#).

3.1 Modelos Aditivos Generalizados para Localização, Escala e Forma (*GAMLSS*)

O modelo aditivo generalizado para Localização, escala e forma (*GAMLSS*) relaxa as suposições de distribuição para a variável resposta de forma que permite modelar a média bem como momentos maiores (escala e forma) como funções das covariáveis. É especialmente útil no caso em que, por exemplo, a resposta não pertence a família exponencial ou quando o interesse está nos parâmetros de escala e forma. Entretanto, os efeitos das covariáveis podem ter formas flexíveis tais como, por exemplo, linear, não linear ou espacial. Portanto, cada parâmetro da distribuição é ligado a um preditor aditivo de forma semelhante ao modelo aditivo generalizado bem estabelecido ([RIGBY; STASINOPOULOS, 2005](#)).

Os termos de um preditor aditivo são geralmente representados por uma abordagem de função de base, o que leva a uma estrutura de modelo genérica e pode ser mais explorada porque cada termo pode ser transformado em uma representação de modelo misto ([RUPPERT; WAND; CARROLL, 2003](#)). Em uma configuração inteiramente Bayesiana, essa generalidade permanece porque as prioris sobre os parâmetros podem também ser formalizadas de uma maneira geral, por exemplo ao designar prioris normais a coeficientes de uma regressão de termos suavizados ([FAHRMEIR *et al.*, 2013](#)).

Os *modelos aditivos* são considerados uma extensão natural dos modelos lineares, os quais mantêm o efeito aditivo na relação entre a variável resposta e as variáveis explicativas podendo incluir funções arbitrárias, não necessariamente lineares, e são

expressos por

$$y_i = \sum_{j=1}^p f_j(x_{ij}) + \varepsilon_i, \quad (3.1)$$

em que $y^\top = (y_1, y_2, \dots, y_n)$ é um vetor $n \times 1$ de respostas e a i -ésima linha da matriz \mathbf{X} é $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ij})$ um vetor $n \times 1$ de variáveis explicativas, com $E(\varepsilon_i) = 0$ e $Var(\varepsilon_i) = \sigma^2$. Note que ao considerar $f_j(x_{ij}) = x_{ij}\beta_j$, tem-se um caso particular do modelo linear.

Este modelo é não paramétrico e cada função f_j , com $j = 1, \dots, p$ é estimada através de suavizadores. Quando os preditores combinam formas paramétricas de algumas variáveis explicativas, (g), com termos não-paramétricos de outras, ($k - g$), são denominados *semiparamétricos* e expressos por

$$y_i = \beta_1 x_{i1} + \dots + \beta_g x_{ig} + f_1(x_{i,g+1}) + \dots + f_{k-g}(x_{i,g+k}) + \varepsilon_i.$$

As funções de suavização, ou suavizadores, de acordo com [Lima, André e Singer \(2001\)](#), são ferramentas que descrevem a variação da média de uma variável Y como função de uma ou mais variáveis não estocásticas X_1, \dots, X_k . Entre os muitos métodos para estimação de modelos não paramétricos ([HäRDLE, 1990](#)), destaca-se Splines.

[Rigby e Stasinopoulos \(2005\)](#) propuseram os modelos aditivos generalizados para locação, escala e forma (GAMLSS), uma nova classe de modelos de regressão (semi)paramétricos, que permitem que todos os parâmetros da variável resposta sejam modelados de forma linear ou não-linear.

A classe de modelos GAMLSS é dita semi-paramétrica, pois exige uma distribuição paramétrica para a variável resposta ao mesmo tempo que permite que os parâmetros da distribuição e das funções das covariáveis sejam modelados através de funções de suavização não-paramétricas. As opções para a distribuição da variável resposta são bastante variadas, existindo cerca de 40 tipos diferentes no GAMLSS, com um, dois, três ou até quatro parâmetros.

Seja uma função de ligação $g_k(\cdot)$, para $k = 1, \dots, p$ que relaciona o k -ésimo parâmetro θ_k às variáveis explanatórias e efeitos aleatórios por meio de um modelo

aditivo dado por:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = X_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk}, \quad (3.2)$$

em que

- $\boldsymbol{\theta}_k^\top = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})$ vetor $n \times 1$;
- $\boldsymbol{\eta}_k^\top = (\eta_{1k}, \eta_{2k}, \dots, \eta_{nk})$ vetor $n \times 1$;
- $\boldsymbol{\beta}_k^\top = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k})$ é um vetor de parâmetros de tamanho J'_k ;
- X_k é uma matriz de especificação conhecida de tamanho $n \times J'_k$;
- Z_{jk} é uma matriz de especificação fixa e conhecida de ordem $n \times q_{jk}$;
- γ_{jk} é uma variável aleatória q_{jk} -dimensional com distribuição $\gamma_{jk} \sim N(0, G_{jk}^{-1})$;
- G_{jk}^{-1} é uma inversa (generalizada) de uma matriz simétrica $G_{jk} = G_{jk}(\boldsymbol{\lambda}_{jk})$ de dimensão $q_{jk} \times q_{jk}$ que pode depender de um vetor de hiperparâmetros $\boldsymbol{\lambda}_{jk}$.

O modelo acima é denotado por *GAMLSS aditivo semiparamétrico*. [Cribari-Neto \(2010\)](#) mostram que como na maioria das situações práticas são requeridos no máximo quatro parâmetros para a distribuição, os quais são caracterizados por $\boldsymbol{\mu}$ (posição), $\boldsymbol{\sigma}$ (escala), \boldsymbol{v} (assimetria) e $\boldsymbol{\tau}$ (curtose) e denominados parâmetros de posição, escala e forma (os dois últimos) respectivamente, para estimá-los temos as seguintes equações:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = X_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} Z_{j1} \gamma_{j1}, \quad (3.3)$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = X_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} Z_{j2} \gamma_{j2}, \quad (3.4)$$

$$g_3(\boldsymbol{v}) = \boldsymbol{\eta}_3 = X_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} Z_{j3} \gamma_{j3}, \quad (3.5)$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = X_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} Z_{j4} \gamma_{j4}. \quad (3.6)$$

GAMs e GAMLSSs estão disponíveis em vários pacotes, principalmente o pacote **mgcv** (WOOD, 2003) e também a família de pacotes **gamlss** (RIGBY; STASINOPOULOS, 2005) e **VGAM** (YEE, 2020). Os dois últimos são notáveis por seu suporte a uma ampla gama de distribuições de resposta. No entanto, para estruturas de previsão complexas e resposta com distribuições além da família exponencial, a estimativa pode ser desafiadora ou sujeita a instabilidades numéricas. Em contraste, **mgcv** se destaca no fornecimento de algoritmos altamente otimizados para modelos suaves gerais, bem como a função **bam()** dedicada para big data.

3.2 Pacote BAMLSS

Esta seção oferece uma visão geral da funcionalidade do pacote *BAMLSS* que em português denominamos Modelos Aditivos Bayesiano para Localização, Escala e Forma.

Umlauf, Klein e Zeileis (2018a) apresentam um conceito unificado, *Lego toolbox*, para modelos de regressão complexos. Os autores mostram que algoritmos de estimação iterativos, por exemplo, para a estimação da média ou moda a posteriori baseado em simulações MCMC, com amostrador JAGS (*Just Another Gibbs Sampler*), exibem estrutura muito semelhantes de tal forma que o processo de construção de modelos se torne simples, já que a arquitetura do modelo é apenas uma combinação de únicos “bricks”. Devido a muitos paralelos com a classe *GAMLSS*, a estrutura conceitual foi chamada de *BAMLSS*.

Baseado nos dados para $i = 1, \dots, n$ observações, o modelo assume independência condicional. Assim como na classe de *GAMLSS* ou modelos de regressão distribucional todos os parâmetros da distribuição podem ser modelados através de covariáveis tais como:

$$y \sim D(h_1(\theta_1) = \eta_1, h_2(\theta_2) = \eta_2, \dots, h_p(\theta_p) = \eta_p), \quad (3.7)$$

em que D denota a distribuição paramétrica para a variável resposta y com p parâmetros θ_k , $k = 1, \dots, p$, que são relacionados aos preditores aditivos usando funções duas vezes diferenciáveis e monotônicas $h_k(\cdot)$. Vale lembrar que a resposta pode ser um vetor q -dimensional $\mathbf{y} = (y_1, \dots, y_q)^\top$, por exemplo, quando D é uma distribuição multivariada.

O k -ésimo preditor aditivo é dado por:

$$\eta_k = \eta_k(\mathbf{x}_k; \beta_k) = f_{1k}(\mathbf{x}; \beta_{1k}) + \cdots + f_{J_k k}(\mathbf{x}; \beta_{J_k k}), \quad (3.8)$$

com funções não especificadas (possivelmente não linear) $f_{jk}(\cdot)$ de subvetores de um vetor \mathbf{x} que coleta toda informação da covariável disponível, $j = 1, \dots, J_k$ e $k = 1, \dots, p$ e $\beta_k = (\beta_{1k}, \dots, \beta_{J_k k})^\top$ são parâmetros, tipicamente coeficientes de regressão que precisam ser estimados a partir dos dados. O vetor de funções de avaliação $\mathbf{f}_{jk} = (f_{jk}(\mathbf{x}_1; \beta_{jk}), \dots, f_{jk}(\mathbf{x}_n; \beta_{jk}))^\top$ de $i = 1, \dots, n$ observações é então dada por

$$f_{jk} = \begin{pmatrix} f_{jk}(x_{1jk}; \beta_{jk}) \\ \vdots \\ f_{jk}(x_{njk}; \beta_{jk}) \end{pmatrix} = f_{jk}(\mathbf{x}_{jk}; \beta_{jk}), \quad (3.9)$$

em que \mathbf{X}_{jk} ($n \times m_{jk}$) é a matriz de especificação e a estrutura de \mathbf{X}_{jk} depende apenas do tipo de covariáveis e das prioris assumidas para $f_{jk}(\cdot)$. Nesta notação o k -ésimo vetor de parâmetros é dado por

$$h_k(\theta_k) = \eta_k = \eta_k(X_k; \beta_k) = f_{1k} + \cdots + f_{J_k k}, \quad (3.10)$$

em que $X_k = (X_{1k}, \dots, X_{J_k k})^\top$ é a matriz de especificação combinada para o k -ésimo parâmetro. As funções $f_{jk}(\cdot)$ são usualmente baseadas na abordagem de uma função base, em que η_k é então um típico tipo GAM ou também como conhecido preditor aditivo estruturado, neste pacote relaxa-se essa suposição e $f_{jk}(\cdot)$ é uma composição não especificada de dados de covariáveis x e coeficientes de regressão β_{jk} ($q_{jk \times 1}$). No caso é derivado de uma abordagem de função de base, o qual pode ser reescrito como

$$f_{jk} = X_{jk} \beta_{jk}, \quad (3.11)$$

no entanto mais geral e termos complexos são permitidos no BAMLSS. Um simples exemplo para uma $f_{jk}(\cdot)$ que é não linear nos parâmetros β_{jk} poderia ser uma curva de crescimento Gompertz

$$f_{jk} = \beta_1 \exp(-\exp(\beta_2 + X_{jk} \beta_3)). \quad (3.12)$$

3.2.1 Resposta e distribuição posteriori

O principal bloco da construção de um algoritmo de modelo de regressão é a função densidade de probabilidade, ou por razões computacionais seu logaritmo. Estimacão tipicamente requer avaliar a função de máxima verossimilhança, ver [Umlauf, Klein e Zeileis \(2018b\)](#)

$$l(\beta; y, X) = \sum_{i=1}^n \log d_y(y_i; \theta_{i1} = h_1^{-1}(\eta_1(x_i; \beta_1)), \dots, \theta_{ik} = h_k^{-1}(\eta_k(x_i; \beta_k))), \quad (3.13)$$

um número razoável de vezes, na qual o vector $\beta = (\beta_1^T, \dots, \beta_k^T)^T$ compreende todos os coeficientes/parâmetros de regressão que devem ser estimados. Determinando distribuições a priori para a componente individual chega-se a log-posteriori

$$\log \pi(\beta, \tau; y, X, \alpha) \propto l(\beta; y, X) + \sum_{k=1}^K \sum_{j=1}^{J_k} [\log p_{jk}(\beta_{jk}; \tau_{jk}, \alpha_{jk})], \quad (3.14)$$

na qual $\tau = (\tau_1^T, \dots, \tau_k^T)^T = (\tau_{11}^T, \dots, \tau_{j_1}^T, \dots, \tau_{1_k}^T, \dots, \tau_{j_k}^T)^T$ é o vetor de todos os hiperparâmetros determinados usados nas funções priori $p_{jk}(\cdot)$ e similarmente α é o conjunto de todas especificações das prioris fixadas. Mais precisamente, a priori mais geral para o j -ésimo termo do modelo do k -ésimo parâmetro é dada por

$$P_{jk}(\beta_{jk}; \tau_{jk}, \alpha_{jk}) \propto p_{\beta_{jk}}(\beta_{jk} | \tau_{jk}; \alpha_{jk}) \cdot p_{\tau_{jk}}(\tau_{jk} | \alpha_{\tau_{jk}}). \quad (3.15)$$

Com densidade priori (ou combinação das densidades) $p_{\beta_{jk}}(\cdot)$ e $p_{\tau_{jk}}(\cdot)$ que depende do tipo da covariável e suposições a priori sobre $f_{jk}(\cdot)$. Nesta estrutura, τ_{jk} são tipicamente variâncias, por exemplo, que explicam o grau de suavidade de $f_{jk}(\cdot)$ ou o valor da correção entre as observações. Por exemplo, usando uma representação spline de $f_{jk}(\cdot)$ em combinação com uma priori normal para $p_{\beta_{jk}}(\cdot)$, as variâncias podem ser interpretadas como o inverso do parâmetro de suavização em um contexto de regressão penalizada, tal que, de uma perspectiva frequentista pode ser vista como uma log-verossimilhança penalizada. Adicionalmente, as especificações fixadas a priori

$\alpha_{jk} = [\alpha_{\beta_{jk}}, \alpha_{\tau_{jk}}]$ pode ainda controlar a forma de $p_{\beta_{jk}}(\cdot)$ e $p_{\tau_{jk}}(\cdot)$ incorporando conhecimentos a priori sobre $\beta_{jk}(\cdot)$ ou para modelos GAM α_{jk} usualmente mantem a então chamada matriz de penalidade, entre outros.

3.2.2 Ajuste do modelo

Estimativas pontuais bayesianas são tipicamente resolvidas através de um dos seguintes métodos:

1. Maximização da log-posteriori para estimação da moda a posteriori
2. Solução de integrais de alta dimensão, por exemplo, para a estimação da média ou moda a posteriori

Para os modelos possivelmente muito complexos com o BAMLSS os problemas 1 e 2 são comumente resolvidos por algoritmos iterativos computacionalmente intensivos, já que as soluções analíticas estão disponíveis apenas em poucos casos especiais. Em ambos os casos, o algoritmo procede com um esquema de atualização do tipo:

$$\left(\beta^{(t+1)}, \tau^{(t+1)}\right)^T = U\left(\beta^{(t)}, \tau^{(t)}; y, X, \alpha\right), \quad (3.16)$$

em que a função $U(\cdot)$ é uma função de atualização, por exemplo, para a geração do primeiro passo do Newton-Raphson do item 1, acima, ou para pegar o próximo passo na simulação MCMC no item 2, entre outras. O esquema de atualização pode ser particionado em equações de atualização separadas usando *leapfrog*, que estima os descendentes a partir dos padrões dos pais, ou iterações zigzag.

Seja:

$$\begin{aligned}
\left(\beta_1^{(t+1)}, \tau_1^{(t+1)}\right)^T &= U_1 \left(\beta_1^{(t)}, \beta_2^{(t)}, \dots, \beta_k^{(t)}, \tau_1^{(t)}, \tau_2^{(t)}, \dots, \tau_k^{(t)}; y, X_1, \alpha_1\right), \\
\left(\beta_2^{(t+1)}, \tau_2^{(t+1)}\right)^T &= U_2 \left(\beta_1^{(t+1)}, \beta_2^{(t)}, \dots, \beta_k^{(t)}, \tau_1^{(t+1)}, \tau_2^{(t)}, \dots, \tau_k^{(t)}; y, X_2, \alpha_2\right), \\
&\vdots \\
\left(\beta_k^{(t+1)}, \tau_k^{(t+1)}\right)^T &= U_k \left(\beta_1^{(t+1)}, \beta_2^{(t+1)}, \dots, \beta_k^{(t)}, \tau_1^{(t+1)}, \tau_2^{(t+1)}, \dots, \tau_k^{(t)}; y, X_k, \alpha_k\right).
\end{aligned} \tag{3.17}$$

O esquema de atualização particionada com função de atualização $U_k(\cdot)$, isto é, em cada iteração de atualização para o k -ésimo parâmetro são obtidos enquanto os demais parâmetros são fixados. Esta estratégia pode ser aplicada para todos os termos dentro de um parâmetro

$$\left(\beta_{jk}^{(t+1)}, \tau_{jk}^{(t+1)}\right)^T = U_{jk} \left(\beta_{jk}^{(t)}, \tau_{jk}^{(t)}, \cdot\right) \quad j = 1, \dots, J_k, k = 1, \dots, K, \tag{3.18}$$

em que $U_{jk}(\cdot)$ é uma função de atualização para um único termo do modelo.

O sistema de atualização particionada permite ter diferentes funções $U_{jk}(\cdot)$ para diferentes termos do modelo, isto é, no item 1 algumas funções de atualização podem ser usadas em mínimos quadrados iterativos ponderados (IWLS) e alguns passos do Newton-Raphson ordinário. No problema do item 2 usando simulação MCMC, é comum misturar vários métodos de amostragem dependendo do tipo do termo do modelo ou da distribuição do parâmetro.

Usando sistemas altamente modulares como (3.17) e (3.18), é possível desenvolver um algoritmo de estimação genérica para inúmeros modelos possivelmente muito complexos. O algoritmo começa iniciando todos os parâmetros do modelo e preditores. Em seguida uma iteração percorre todas as performances dos parâmetros das distribuições e uma iteração interna atualiza todos os termos do modelo do respectivo parâmetro, isto é, o algoritmo usa um esquema de atualização do tipo *backfitting*. Na prática, para inferência Bayesiana completa o algoritmo é aplicado duas vezes, isto é, primeiro calcula estimativas para (3.17) e então usando estas estimativas como ponto de partida completa o processo resolvendo (3.18).

Encontrar bons valores iniciais é especialmente importante para modelos complexos, por exemplo, para funções multidimensionadas $f_{jk}(\cdot)$ que tem múltiplas suavizações da variância na densidade a priori $p_{jk}(\cdot)$. Portanto propomos estimar os parâmetros τ_{jk} usando um critério de qualidade de ajuste na abordagem de seleção via stepwise.

Em cada passo da atualização, cada $\tau_{jk} = (\tau_{1jk}, \dots, \tau_{Ljk})^T$ é otimizado um após o outro usando intervalos de busca adaptativos. Assim, o problema de otimização é reduzido a uma busca unidimensional que é relativamente rápida e simples de implementar. O algoritmo não garante um mínimo global dado o critério de qualidade do ajuste, no entanto, a solução é pelo menos perto e serve como bom ponto inicial para o MCMC. A velocidade da otimização pode ser posteriormente aumentada se para um intervalo de busca dado somente um gride de valores possíveis para cada τ_{ljk} é usado.

As funções atualizadas do MCMC usualmente ou aceita ou rejeita amostras dos parâmetros e variâncias suavizadas são amostradas depois de β_{jk} . A estrutura geral de amostragem do algoritmo, ou seja, a função proposta gera amostra dos parâmetros $\beta_{jk}^{t+1}, \tau_{jk}^{t+1}$ usando, possivelmente, diferentes esquemas de amostragem como: baseados no Metropolis-Hastings e amostrador *slice*.

3.2.3 Lego Bricks

Para obtermos os parâmetros atualizados para os itens 1 ou 2 usando sistemas de atualizações particionadas flexíveis como 3.17 e 3.18 o seguinte "Lego bricks" é repetidamente usado no algoritmo:

1. A densidade $d_y(y|\theta_1, \dots, \theta_k)$ e sua respectiva função de log-verossimilhança $l(\beta; y, X)$,
2. Função de ligação $h_k(\cdot)$,
3. Os termos do modelo $f_{jk}(\cdot)$ e suas correspondentes densidades a priori $p_{jk}(\beta_{jk}; \tau_{jk}, \alpha_{jk})$.

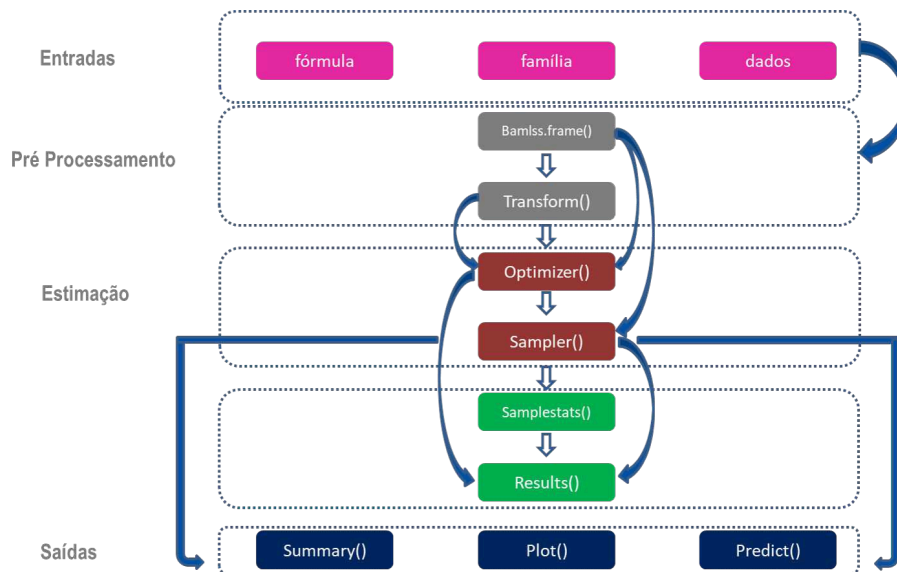
Mais ainda, também apresentamos para atualização do algoritmo que usualmente requer:

1. A derivada da inversa da função de ligação $h_k^{-1}(\cdot)$,

2. A derivada de primeira ordem dos preditores $\frac{\partial \eta_k}{\partial \beta_{jk}}$,
3. A derivada de primeira ordem da função log-verossimilhança com respeito aos coeficientes/parâmetros de regressão $\frac{\partial l(\beta; y, X)}{\partial \beta_{jk}}$ com respeito às preditoras $\frac{\partial l(\beta; y, X)}{\partial \eta_k}$
4. A derivada de segunda ordem da função log-verossimilhança com respeito aos coeficientes/parâmetros de regressão $\frac{\partial^2 l(\beta; y, X)}{\partial \beta_{jk} \partial \beta_{jk}}$ com respeito às preditoras $\frac{\partial^2 l(\beta; y, X)}{\partial \eta_k \partial \eta_k}$
5. A derivada das log-prioris, isto é, $\frac{\partial p_{jk}(\beta_{jk}; \tau_{jk}, \alpha_{jk})}{\partial \beta_{jk}}$.

A Figura 3.2.1 evidencia o funcionamento do pacote:

Figura 3.2.1 – Funcionamento do pacote.



Fonte: Elaborada pelo autor.

3.2.4 Exemplo de uso do pacote

Nesta subseção, apresentaremos um exemplo de uso do pacote BAMLSS desde sua instalação até aplicação em um conjunto de dados. Nos baseamos no artigo (UMLAUF *et al.*, 2019) que apresenta bons exemplos sobre o uso do pacote.

Para o exemplo, faremos uso dos dados retirados do pacote AER do *software R* sobre participação das mulheres na força de trabalho na Suécia em 1981. A base de dados contém 872 observações de 6 variáveis, algumas podem ter uma influência não linear na resposta (participação no trabalho). Para esses dados, um modelo Bayesiano Logito binomial usando o algoritmo MCMC pode ser ajustado. Mostraremos os passos iniciais para modelagem, bem como outras etapas de uso do pacote que podem facilitar entendimento dos dados e como modelar fazendo uso deste pacote.

Para tanto repetiremos aqui como é feita a construção do modelo e descreveremos as funções que precisaremos para declarar no pacote. Desta maneira temos que:

- $y \sim D(h_k(\theta))$ caracteriza nossa fórmula que é uma função linear entre as covariáveis e a resposta
- D denota a distribuição para a resposta y
- h_k é uma função duplamente diferenciável e monotônica (ex. distribuições da família exponencial)

Para usar como exemplo de aplicação temos os dados "SwissLabor" do pacote AER. Os dados são referentes a dados de um corte transversal originado da pesquisa de saúde "SOMIPOPS" para a suíça em 1981. Contendo 872 observações e 7 atributos no total, sendo eles:

- **Participation:** Variável que indica se o indivíduo participa da força de trabalho;
- **Income:** Logaritmo da receita não-laboral;
- **Age:** Idade em décadas (Anos divididos por 10);
- **Education:** Anos de formação educacional;

- **YoungKids:** Número de crianças jovens (abaixo de 7 anos de idade);
- **OldKids:** Número de crianças mais velhas (acima de 7 anos de idade);
- **Foreign:** Variável que indica se o indivíduo é estrangeiro (isto é: não Suíço)

Sendo assim, abaixo apresentamos os passos de como usar esse modelo dentro do pacote BAMLSS no R.

```
1 #Instalando o pacote:
2 install.packages('bamlss')
3
4 #Carrega o pacote:
5 library("bamlss")
6
7 #Carregando os dados:
8 data("SwissLabor", package= "AER")
9
10 #Fórmula do modelo a ser ajustado
11 f = participation ~ income + age + education + yougkids +
12     oldkids + foreign + I(age^2)
13
14 #definição da semente:
15 set.seed(123)
16
17 #Ajuste do modelo:
18 b = bamlss(f, family = "binomial", data=SwissLabor)
19
20 #Obtendo os resultados:
21 summary(b)
```

Código-fonte 1 – Código fonte em R - Configuração inicial e uso do BAMLSS

Até este passo, temos algumas informações adicionais a serem acrescentadas. Por padrão o número de iterações usadas no amostrador MCMC é 1200, o aquecimento é de 200 e o espaçamento, ou salto, é 1. A razão é que durante o processo de modelagem usualmente queremos obter os primeiros resultados mais rapidamente. O ajuste padrão do

modelo usa objetos similares aos que são usados na função *glm* que podem ser facilmente implementados para novas distribuições (modelos).

No exemplo mostrado, note que para capturar não linearidades, um termo quadrático para a variável *age* foi adicionado ao modelo. O objeto resultante em *b* é uma classe "bamlss" que por padrão podemos aplicar funções como: *summary()*, *coef()*, *plot()*, *predict()*, etc.

Já nos comandos a seguir, apresentamos uma forma de observar o comportamento daqueles que participam da força de trabalho e aqueles que não participam a fim de entender, por exemplo, como função das idades.

```
1 dados = read.table(file.choose(), header=T)
2 attach(dados)
3 summary(dados)
4 install.packages("bamlss")
5
6 ### Gerfin (1996), Table I.
7 fm_probit <- glm(LFP ~ . + I(AGE^2), data = dados,
8                 family = binomial(link = "probit"))
9
10 summary(fm_probit)
11
12 set.seed(123)
13 f = LFP ~ LNNLINC+AGE+EDUC+NYC+NOC+FOREIGN + I(AGE^2)
14 b = bamlss(f, family = "binomial", data = dados)
15
16 summary(b)
17 plot(b, which = c("samples", "max-acf"))
18
19
20 nd <- data.frame(LNNLINC = 11, AGE = seq(2, 6.2, length =
21               100),
22               EDUC = 12, NYC = 1, NOC = 1, FOREIGN = 0)
23 nd2 <- data.frame(LNNLINC = 11, AGE = seq(2, 6.2, length =
24               100),
25               EDUC = 12, NYC = 1, NOC = 1, FOREIGN = 1)
26
27 str(nd)
28 str(nd2)
29
30 nd$pSwiss <- predict(b, newdata = nd, type = "parameter",
31                   FUN = c95)
32 nd2$FOREIGN <- 1
33 nd2$pForeign <- predict(b, newdata = nd2, type = "parameter"
34                       , FUN = c95)
```



```

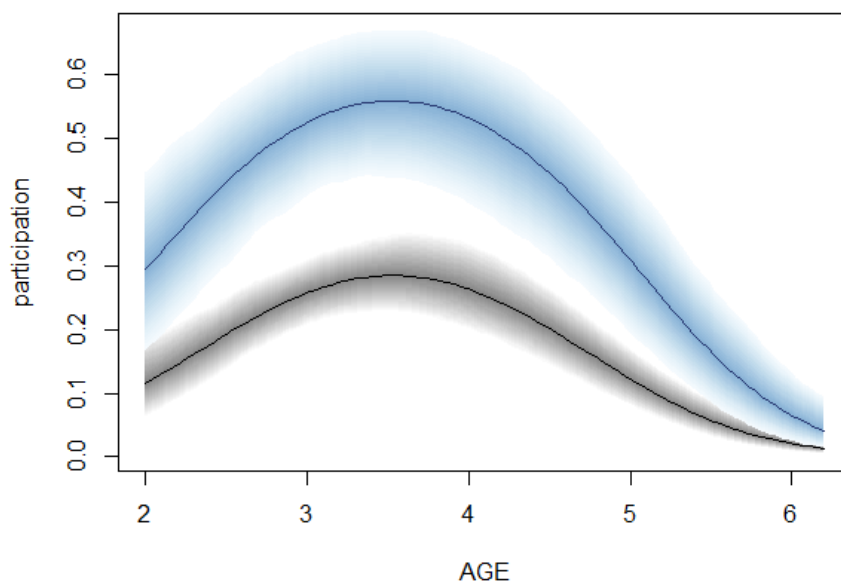
31
32 #Age
33 blues <- function(n, ...) sequential_hcl(n, "Blues", rev =
      TRUE)
34
35 plot2d(pSwiss ~ AGE, data = nd, ylab = "participation",
36       ylim = range(c(nd$pSwiss, nd2$pForeign)),
37       fill.select = c(0, 1, 0, 1))
38 plot2d(pForeign ~ AGE, data = nd2, add = TRUE,
39       fill.select = c(0, 1, 0, 1), axes = FALSE,
40       s2.col = blues, col.lines = blues(1))
41
42 #exemplo 2
43 f2 <- LFP ~ LNNLINC+EDUC+NYC+NOC+FOREIGN + s(AGE, k = 10)
44 set.seed(123)
45 b2 <- bamlss(f2, family = "binomial", data = dados)
46 plot(b2)
47
48 dados$Ag <- cut(dados$AGE,
49               breaks = quantile(dados$AGE, prob = seq(0,
50               1, length = 10)),
51               include.lowest = TRUE, ordered_result = TRUE
52               )
53 f3 <- LFP ~ LNNLINC+EDUC+NYC+NOC+FOREIGN + la(Ag, fuse = 2)
54 b3 <- bamlss(f3, family = "binomial", data = dados,
55             optimizer = lasso, sampler = FALSE,
56             criterion = "BIC", upper = exp(5), lower = 1)
57 pathplot(b3, which = "loglik.contrib", intercept = FALSE)
58 page <- predict(b3, term = "Ag", intercept = FALSE,
59               mstop = lasso_stop(b3))
60 plot2d(page ~ AGE, data = dados, rug = TRUE)

```

Código-fonte 2 – Código fonte em R - Configuração inicial

Na Figura 3.2.2 observamos que, independente da idade, os estrangeiros (curva azul) tem maior força de trabalho do que os não estrangeiros (curva preta). Sendo

Figura 3.2.2 – Comparação entre os participantes da força de trabalho e os não participantes variando a idade, curva azul (extrangeiros) e curva preta (não estrangeiros).



Fonte: Elaborada pelo autor.

que a medida que a idade aumenta essa diferença na participação diminui bastante, ou, trabalhadores estrangeiros tem maior participação na força de trabalho do que os trabalhadores não estrangeiros.

O MODELO DE REGRESSÃO ZARBS

Neste capítulo será estudado o modelo ZARBS proposto por [Tomazella et al. \(2019\)](#). Esta metodologia considera um modelo misto com uma parte discreta que comporta a proporção de zeros e a parte contínua que tem como distribuição a RBS.

4.1 Modelo ZARBS

Segundo [Tomazella et al. \(2019\)](#) dizemos que uma variável aleatória Y possui distribuição ZARBS com parâmetros μ , σ e v , denotada por $ZARBS(\mu, \sigma, v)$ se sua respectiva fdp é dada por:

$$f_Y(y|\mu, \sigma, v) = \left((1-v) \frac{\exp\left(\frac{\sigma}{2}\right) \sqrt{\sigma+1}}{4y^{\frac{3}{2}} \sqrt{\pi\mu}} \left[y + \frac{\sigma\mu}{\sigma+1} \right] \right) \times \exp\left(-\frac{\sigma}{4} \left[\frac{(\sigma+1)y}{\sigma\mu} + \frac{\sigma\mu}{(\sigma+1)y} \right] \right)^{1-I(y=0)} \times v^{I(y=0)}, \quad (4.1)$$

em que $y > 0$, $\sigma > 0$, $\mu > 0$ e $0 < v < 1$ e $I(\cdot)$ representa a função indicadora. Como resultados temos que $\mathbb{E}[Y] = (1-v)\mu$ e $Var[Y] = (1-v)\mu^2 \times \left[v + CV(T)^2 \right]$ na qual $CV = \frac{\sqrt{(2\sigma+5)}}{\delta+1}$. A partir daí podemos obter a função ln-FDP dado por:

$$l(\mu, \sigma, v) = \ln(v)I(y=0) + \ln(1-v)[1-I(y=0)] + [1-I(y=0)] \ln(f(y)). \quad (4.2)$$

Sendo $Y \sim \text{ZARBS}(\mu, \sigma, \nu)$ então a função de máxima verossimilhança de $\theta [\mu, \sigma, \nu]^T$ é dada por:

$$L(\mu, \sigma, \nu) = \prod_{i=1}^n f_y(y_i | \mu, \sigma, \nu) = \nu^{n_0} [1 - \nu]^{n - n_0} \prod_{i=1}^n f(y_i | \mu, \sigma)^{1 - I(y_i=0)}, \quad (4.3)$$

em que $n_0 = \sum_{i=1}^n I(y_i = 0)$ denota a quantidade de observações iguais a zeros na amostra.

Desta maneira, podemos decompor (4.3) em: $l(\theta) = l(\nu) + l(\sigma, \mu)$ e assim podemos separar a verossimilhança em duas partes, expressas como:

$$\begin{aligned} l(\nu) &= n_0 \log(\nu) + [n - n_0] \log(1 - \nu), \\ l(\sigma, \mu) &= [n - n_0] c(\mu, \sigma) - \frac{3}{2} \sum_{y_i > 0} \log(y_i) - \frac{[\delta + 1]}{4\mu} \sum_{y_i > 0} y_i - \sum_{y_i > 0} \frac{\mu \sigma^2}{4[\sigma + 1]y_i} \\ &\quad + \sum_{y_i > 0} \log\left(y_i + \frac{\mu \sigma}{[\delta + 1]}\right), \end{aligned} \quad (4.4)$$

em que $c(\mu, \sigma) = -\frac{1}{2} \log(16\pi) + \frac{\sigma}{2} - \frac{1}{2} \log(\mu) + \frac{1}{2} \log(\sigma + 1)$. Agora considere que Y_1, \dots, Y_n , iid, uma variável aleatória tal que $Y_i \sim \text{ZARBS}(\mu_i, \sigma_i, \nu_i)$ e suponha ainda que os parâmetros que representam a média, a variância e a mistura de Y_i satisfaçam as relações funcionais:

$$\begin{aligned} g_1(\mu_i) &= \eta_i = f_1(X_i; \beta) \\ g_2(\sigma_i) &= \tau_i = f_2(Z_i; \alpha) \\ g_3(\nu_i) &= \varepsilon_i = f_3(W_i; \gamma) \end{aligned} \quad (4.5)$$

Para $i = 1, \dots, n$, $\beta = [\beta_1, \dots, \beta_p]^T$, $\alpha = [\alpha_1, \dots, \alpha_q]^T$ e $\gamma = [\gamma_1, \dots, \gamma_r]^T$ representam os vetores de parâmetros desconhecidos a serem estimados. Ainda temos que $p + q + r < n$, $\eta = [\eta_1, \dots, \eta_n]^T$, $\tau = [\tau_1, \dots, \tau_n]^T$ e $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T$ representam os vetores de preditores e $f_j(\cdot, \cdot)$, $j = 1, 2, 3$ são funções lineares ou não lineares contínuas duas vezes diferenciáveis nos seus segundos argumentos, de forma que suas matrizes de derivadas são $\tilde{X} = \frac{\partial \eta}{\partial \beta}$, $\tilde{Z} = \frac{\partial \tau}{\partial \alpha}$, $\tilde{W} = \frac{\partial \varepsilon}{\partial \gamma}$ de rank completo para β , α , γ . Além do mais $\tilde{x}_i = [\tilde{x}_{i1}, \dots, \tilde{x}_{ip_1}]^T$, $\tilde{z}_i = [\tilde{z}_{i1}, \dots, \tilde{z}_{ip_2}]^T$ e $\tilde{w}_i = [\tilde{w}_{i1}, \dots, \tilde{w}_{ip_3}]^T$ são os vetores de valores p , q e r das respectivas variáveis explicativas.

Além disso as funções $g_j : \mathbb{R}^+ \rightarrow \mathbb{R}$, $j = 1, 2$ são monotonas, positivas e pelo menos duas vezes diferenciáveis enquanto que $g_3 : (0, 1) \rightarrow \mathbb{R}$ é estritamente monotônica

e, ao menos, duas vezes diferenciável. Neste cenário, $\mu_i = g_1^{-1}(\eta_i)$, $\sigma_i = g_1^{-1}(\tau_i)$ e $v_i = g_1^{-1}(\varepsilon_i)$. Vale lembrar também que a primeira expressão dada em (4.4) representa a função de máxima verossimilhança de um modelo com resposta binária enquanto que a segunda expressão representa a função de máxima verossimilhança para $[\beta, \alpha]^T$ em um modelo de regressão não-linear RBS com parâmetro de precisão.

4.1.1 Vetor Score, Matriz Hessiana e de Informação de Fisher

Nesta subsecção, mostraremos como obter as informações Vetor Score, Matriz Hessiana e a Matriz de informação de Fisher. Considere os elementos do vetor de score obtidos da seguinte forma:

$$U_{\beta_j} = \sum_{i=1}^n d_{\mu}^{(i)} a_i \tilde{x}_{ij}, \quad U_{\alpha_r} = \sum_{i=1}^n d_{\sigma}^{(i)} b_i \tilde{z}_{ir}, \quad \text{e} \quad U_{\gamma_j} = \sum_{i=1}^n d_v^{(i)} c_i \tilde{w}_{ij}. \quad (4.6)$$

Sabendo que $\ell(\mu_i, \sigma_i, v_i)$ é da maneira como mostrada em (4.4) podemos obter as derivadas com respeito a cada parâmetro onde para $j = 1, \dots, p$, $r = 1, \dots, q$ e $j = 1, \dots, r$ as derivadas são dadas por:

$$\begin{aligned} d_{\mu}^{(i)} &= \underbrace{[1 - \mathbb{I}(y_i = 0)]}_{\kappa_i} \left\{ \underbrace{\frac{[\sigma_i + 1]}{4\mu_i^2} - \frac{\sigma_i^2}{4[\sigma_i + 1]} \frac{1}{y_i} + \frac{\sigma_i}{[\sigma_i + 1]} \frac{1}{\left[y_i + \frac{\mu_i \sigma_i}{\sigma_i + 1}\right]}}_{y_i^*} - \underbrace{\frac{1}{2\mu_i}}_{\mu_i^*} \right\} \\ &= \kappa_i (y_i^* - \mu_i^*), \\ d_{\sigma}^{(i)} &= \underbrace{[1 - \mathbb{I}(y_i = 0)]}_{\kappa_i} \left\{ \underbrace{\frac{\mu_i}{[\sigma_i + 1]^2} \frac{1}{\left[y_i + \frac{\mu_i \sigma_i}{\sigma_i + 1}\right]} - \frac{1}{4\mu_i} y_i - \frac{\mu_i \sigma_i [\sigma_i + 2]}{4[\sigma_i + 1]^2} \frac{1}{y_i} + \frac{[\sigma_i + 2]}{2[\sigma_i + 1]}}_{y_i^{\bullet}}}_{-\sigma_i^{\bullet}} \right\} \\ &= \kappa_i (y_i^{\bullet} - \sigma_i^{\bullet}), \\ d_v^{(i)} &= \frac{1}{v_i} \mathbb{I}(y_i = 0) - \frac{1}{1 - v_i} [1 - \mathbb{I}(y_i = 0)] = \underbrace{\frac{1}{v_i [1 - v_i]} \mathbb{I}(y_i = 0)}_{y_i^{\circ}} - \underbrace{\frac{1}{[1 - v_i]}}_{v_i^{\circ}} = (y_i^{\circ} - v_i^{\circ}), \end{aligned}$$

em que $a_i = \frac{1}{g'_1(\mu_i)}$, $b_i = \frac{1}{g'_2(\sigma_i)}$ e $c_i = \frac{1}{g'_3(v_i)}$. No entanto, a partir de (4.6) e (4.7), reescrevemos o vetor score $(p+q+r) \times 1 U_\theta$ como $[U_\beta^\top, U_\alpha^\top, U_\gamma^\top]^\top$ na qual $U_\beta = \tilde{X}^\top A_\kappa [y^* - \mu^*]$, $U_\alpha = \tilde{Z}^\top B_\kappa [y^\bullet - \sigma^\bullet]$ e $U_\gamma = \tilde{W}^\top C [y^\circ - v^\circ]$, na qual $A_\kappa = \kappa_i a_i \delta_{ij}^n$, $B_\kappa = \kappa_i b_i \delta_{ij}^n$ e $C = c_i \delta_{ij}^n$, onde δ_{ij}^n é o delta de Kronecker. Desta maneira, A_κ , B_κ e C são matrizes diagonal $n \times n$. Além disso, temos que $\tilde{X} = \partial \eta / \partial \beta$, $\tilde{Z} = \partial \tau / \partial \alpha$ e $\tilde{W} = \partial \xi / \partial \gamma$ são matrizes de derivadas com rank p , q e r , respectivamente, para todo β , α e γ . A matriz Hessiana é dada por:

$$h_\theta = \begin{bmatrix} \tilde{X}^\top D \tilde{X} & \tilde{X}^\top M \tilde{Z} & 0 \\ \tilde{Z}^\top M \tilde{X} & \tilde{Z}^\top E \tilde{Z} & 0 \\ 0 & 0 & \tilde{W}^\top O \tilde{W} \end{bmatrix}. \quad (4.7)$$

Os elementos da matriz bloco h_θ observada em (4.7), pode ser obtida através das derivadas parciais. Para $j = 1, \dots, p$ e $r = 1, \dots, q$, chegamos que

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \beta_{j'}} &= \sum_{i=1}^n \underbrace{[d_{\mu^2}^{(i)} a_i + d_{\mu'}^{(i)} a'_i]}_{d_i} a_i \tilde{x}_{ij} \tilde{x}_{i k}, & \frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \alpha_R} &= \sum_{i=1}^n \underbrace{d_{\mu\sigma}^{(i)} a_i b_i}_{m_i} \tilde{x}_{ij} \tilde{z}_{i k}, & \frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \gamma_j} &= 0, \\ \frac{\partial^2 \ell(\theta)}{\partial \alpha_R \partial \alpha_{R'}} &= \sum_{i=1}^n \underbrace{[d_{\sigma^2}^{(i)} b_i + d_{\sigma'}^{(i)} b'_i]}_{e_i} b_i \tilde{z}_{ij} \tilde{z}_{i k}, & \frac{\partial^2 \ell(\theta)}{\partial \alpha_R \partial \gamma_j} &= 0, & \frac{\partial^2 \ell(\theta)}{\partial \gamma_j \partial \gamma_{j'}} &= \sum_{i=1}^n \underbrace{[d_{v^2}^{(i)} c_i + d_{v'}^{(i)} c'_i]}_{o_i} c_i \tilde{w}_{ij} \tilde{w}_{i k}, \end{aligned} \quad (4.8)$$

na qual

$$\begin{aligned} d_{\mu^2}^{(i)} &= \kappa_i \left\{ \frac{1}{2\mu_i^2} - \frac{\sigma_i^2}{[\sigma_i y_i + y_i + \sigma_i \mu_i]^2} - \frac{y_i [\sigma_i + 1]}{2\mu_i^3} \right\}, \\ d_{\mu v}^{(i)} &= 0, \\ d_{\mu\sigma}^{(i)} &= \kappa_i \left\{ \frac{y_i}{[\sigma_i y_i + y_i + \sigma_i \mu_i]^2} + \frac{y_i}{4\mu_i^2} - \frac{\sigma_i [\sigma_i + 2]}{4[\sigma_i + 1]^2 y_i} \right\}, \\ d_{\sigma v}^{(i)} &= 0, \\ d_{\sigma^2}^{(i)} &= \kappa_i \left\{ \frac{1}{2[\sigma_i + 1]^2} - \frac{[y_i + \mu_i]^2}{[\sigma_i y_i + y_i + \sigma_i \mu_i]^2} - \frac{\mu_i}{2[\sigma_i + 1]^3 y_i} \right\}, \\ d_{v^2}^{(i)} &= \frac{2v_i - 1}{v_i^2 [1 - v_i]^2} \mathbb{I}(y_i = 0) - \frac{1}{[1 - v_i]^2}. \end{aligned} \quad (4.9)$$

A partir de (4.8) e (4.9) podemos chegar a seguinte matriz $\mathbf{h}_\beta = \tilde{\mathbf{X}}^\top \mathbf{D}, \tilde{\mathbf{X}}, \mathbf{h}_{\beta\alpha} = \tilde{\mathbf{X}}^\top \mathbf{M}, \tilde{\mathbf{Z}}, \mathbf{h}_\alpha = \tilde{\mathbf{Z}}^\top, \mathbf{E}, \tilde{\mathbf{Z}}$ e $\mathbf{h}_\gamma = \tilde{\mathbf{W}}^\top \mathbf{O} \tilde{\mathbf{W}}$, sendo $\mathbf{M} = m_i \delta_{ij}^n$, $\mathbf{D} = d_i \delta_{ij}^n$, $\mathbf{E} = e_i \delta_{ij}^n$ e $\mathbf{O} = o_i \delta_{ij}^n$. Assim, a matriz de informação de Fisher é dada por:

$$\mathbf{i}_\theta = \begin{bmatrix} \tilde{\mathbf{X}}^\top \mathbf{V} \tilde{\mathbf{X}} & \tilde{\mathbf{X}}^\top \mathbf{S} \tilde{\mathbf{Z}} & \mathbf{0} \\ \tilde{\mathbf{Z}}^\top \mathbf{S} \tilde{\mathbf{X}} & \tilde{\mathbf{Z}}^\top \mathbf{U} \tilde{\mathbf{Z}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{W}}^\top \mathbf{Q} \tilde{\mathbf{W}} \end{bmatrix}, \quad (4.10)$$

em que $\mathbf{V} = E_{d_i} \delta_{ij}^n = \text{diag}(E_{d_1}, \dots, E_{d_n})$, $\mathbf{S} = E_{m_i} \delta_{ij}^n = \text{diag}(E_{m_1}, \dots, E_{m_n})$, $\mathbf{U} = E_{e_i} \delta_{ij}^n = \text{diag}(E_{e_1}, \dots, E_{e_n})$ e $\mathbf{Q} = E_{o_i} \delta_{ij}^n = \text{diag}(E_{o_1}, \dots, E_{o_n})$. Sendo

$$\begin{aligned} E_{d_i} &= -E[d_i] = [1 - v_i] \left\{ \frac{\sigma_i}{2\mu_i^2} + \frac{\sigma_i^2}{[\sigma_i + 1]^2} \lambda_i \right\} a_i^2, \\ \mathcal{E}_{m_i} &= -\mathbb{E}[m_i] = [1 - v_i] \left\{ \frac{1}{2\mu_i[\sigma_i + 1]} + \frac{\sigma_i \mu_i}{[\sigma_i + 1]^3} \lambda_i \right\} a_i b_i, \\ \mathcal{E}_{e_i} &= -\mathbb{E}[e_i] = [1 - v_i] \left\{ \frac{[\sigma_i^2 + 3\sigma_i + 1]}{2\sigma_i^2[\sigma_i + 1]^2} + \frac{\mu_i^2}{[\sigma_i + 1]^4} \lambda_i \right\} b_i^2, \\ \mathcal{E}_{o_i} &= -\mathbb{E}[o_i] = \frac{1}{v_i[1 - v_i]}, \end{aligned} \quad (4.11)$$

e

$$\lambda_i = \int_0^\infty \frac{\sqrt{\sigma_i + 1} \exp\{\sigma_i/2\}}{4\sqrt{\pi\mu_i} y_i^{3/2}} \left[y_i + \frac{\sigma_i \mu_i}{\sigma_i + 1} \right]^{-1} \times \exp\left(-\frac{\sigma_i}{4} \left[\frac{[\sigma_i + 1] y_i}{\sigma_i \mu_i} + \frac{\sigma_i \mu_i}{[\sigma_i + 1] y_i} \right] \right) dy_i.$$

A integral λ_i é mais convenientemente calculada a partir da função integrate no software R. A inversa da matriz de informação de Fisher dada em (4.10), é escrita

como:

$$\mathbf{i}_\theta^{-1} = \begin{bmatrix} [\tilde{\mathbf{X}}^\top \mathbf{W}_1 \tilde{\mathbf{X}}]^{-1} & -[\tilde{\mathbf{X}}^\top \mathbf{W}_1 \tilde{\mathbf{X}}]^{-1} \tilde{\mathbf{X}}^\top \mathbf{S} \tilde{\mathbf{Z}} [\tilde{\mathbf{Z}}^\top \mathbf{U} \tilde{\mathbf{Z}}]^{-1} & \mathbf{0} \\ -[[\tilde{\mathbf{X}}^\top \mathbf{W}_1 \tilde{\mathbf{X}}]^{-1} \tilde{\mathbf{X}}^\top \mathbf{S} \tilde{\mathbf{Z}} [\tilde{\mathbf{Z}}^\top \mathbf{U} \tilde{\mathbf{Z}}]^{-1}]^\top & [\tilde{\mathbf{Z}}^\top \mathbf{W}_2 \tilde{\mathbf{Z}}]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & [\tilde{\mathbf{W}}^\top \mathbf{Q} \tilde{\mathbf{W}}]^{-1} \end{bmatrix}, \quad (4.12)$$

em que $\mathbf{W}_1 = \mathbf{V} - \mathbf{S} \tilde{\mathbf{Z}} [\tilde{\mathbf{Z}}^\top \mathbf{U} \tilde{\mathbf{Z}}]^{-1} \tilde{\mathbf{Z}}^\top \mathbf{S}$ e $\mathbf{W}_2 = \mathbf{U} - \mathbf{S} \tilde{\mathbf{X}} [\tilde{\mathbf{X}}^\top \mathbf{V} \tilde{\mathbf{X}}]^{-1} \tilde{\mathbf{X}}^\top \mathbf{S}$. Com as matrizes aumentadas, $\tilde{\mathbf{W}}$, $\tilde{\mathbf{X}}$ e $\tilde{\mathbf{D}}$,

$$\mathbf{W}_{3n,3n} = \begin{bmatrix} \mathbf{V} \mathbf{S} \mathbf{0} \\ \mathbf{S} \mathbf{U} \mathbf{0} \\ \mathbf{0} \mathbf{0} \mathbf{Q} \end{bmatrix}, \quad \tilde{\mathbf{X}}_{3n,p'+q+r} = \begin{bmatrix} \tilde{\mathbf{X}} \mathbf{0} \mathbf{0} \\ \mathbf{0} \tilde{\mathbf{Z}} \mathbf{0} \\ \mathbf{0} \mathbf{0} \tilde{\mathbf{W}} \end{bmatrix}, \quad \text{e} \quad \mathbf{D}_{3n,3n} = \begin{bmatrix} \mathbf{A}^{(m)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{(m)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}^{(m)} \end{bmatrix}$$

A matriz de informação de Fisher pode ser reescrita como $\mathbf{i}_\theta = \mathbf{X}^\top \mathbf{W} \mathbf{X}$ bem como sua inversa $\mathbf{i}_\theta^{-1} = [\mathbf{X}^\top \mathbf{W} \mathbf{X}]^{-1}$. Usando o procedimento iterativo do Score de Fisher, o algoritmo de estimação para $\theta = [\beta^\top, \alpha^\top, \gamma^\top]^\top$ é dado por $\theta^{(m+1)} = \theta^{(m)} + \mathbf{i}_\theta^{-1} U_\theta^{(m)} = [\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \tilde{\mathbf{z}}^{(m)}$ no qual $\tilde{\mathbf{z}}^{(m)} = \mathbf{X} \theta^{(m)} + \{\mathbf{W}^{(m)}\}^{-1} \mathbf{D}^{(m)} \mathbf{z}^{(m)}$ e $\mathbf{z}^{(m)} = [[\mathbf{y}^* - \boldsymbol{\mu}^*]^{(m)\top}, [\mathbf{y}^\bullet - \boldsymbol{\delta}^\bullet]^{(m)\top}, [\mathbf{y}^\circ - \mathbf{v}^\circ]^{(m)\top}]^\top$.

4.2 Análise de Diagnóstico

Nesta seção, apresentaremos um esquema para análise residual e de perturbação de forma a identificar as observações que exercem influência incomum no processo de estimativa e, conseqüente, impacto nas estimativas de parâmetros.

4.2.1 Resíduos Quantílicos

Uma das melhores maneiras de se avaliar a qualidade do ajuste de um modelo partindo das suposições do modelo de regressão ZARBS, foi proposto o resíduo quantílico

aleatorizado (DUNN; SMYTH, 1996) que é uma versão aleatorizada do resíduo proposta por Cox e Snell (1968) e é obtida por:

$$r_i^q = \begin{cases} \Phi^{-1}(f_i), & \text{if } y_i > 0; \\ \Phi^{-1}(u_i), & \text{if } y_i = 0, \end{cases} \quad (4.13)$$

em que $\Phi^{-1}(\cdot)$ é a função inversa da FDA da normal padrão, $f_i = F(y_i, \hat{\mu}_i, \hat{\sigma}_i, \hat{\nu}_i)$ por sua vez é o valor estimado da FDA da distribuição ZARBS e u_i é o valor observado da variável aleatória U_i com distribuição Uniforme no intervalo $(0, \hat{\nu}_i)$, na qual $\hat{\nu} = \hat{P}(Y_i = 0)$. Considerando que o modelo está especificado corretamente então r_i^q tem distribuição aproximadamente normal.

A saída padrão de diagnóstico nos modelos GAMLSS bem como no modelo BAMLSS é composta pelos gráficos de resíduos vs valores ajustados, resíduos vs um índice ou uma variável explicativa e o gráfico quantil-quantil normal (QQ-Plot) dos resíduos.

Uma tendência no gráfico de resíduos contra as preditoras, ou contra os valores preditos, pode indicar uma falta de especificação na função de ligação. Além disso, o gráfico de envelope simulado pode ser um instrumento importante para diagnóstico do modelo (ATKINSON, 1985).

4.2.2 Influência Local

Cook (1986) introduziu o conceito de influência local a qual é baseada na curvatura do plano da função log-verossimilhança por meio da análise de observações perturbadas. O vetor de perturbação $\omega = [\omega_1, \dots, \omega_n]^\top$, $\omega_i \geq 0$, modifica a contribuição de cada caso na log-verossimilhança. A influência de uma observação pode ser avaliada a partir do impacto a que esta observação perturbada, ω , causa na função de log-verossimilhança. A influência local é definida por

$$LD = 2 \left[\ell(\hat{\theta}) - \ell(\hat{\theta}_\omega) \right],$$

na qual $\ell(\hat{\theta})$ é o log da função de verossimilhança do modelo de regressão sem perturbação para o estimador de verossimilhança, $\hat{\theta}$, e $\ell(\hat{\theta}_\omega)$ é o log da função de verossimilhança do modelo de regressão perturbado para o estimador de verossimilhança, $\hat{\theta}_\omega$.

A curvatura normal de LD para θ na direção do vetor d , com $\|d\| = 1$, é dada como $C_d = 2|\mathbf{d}^\top \Delta^\top h_\theta^{-1} \Delta \mathbf{d}|$, na qual $\Delta = \partial^2 \ell(\theta|\omega) / \partial \theta \partial \omega$ é a matriz que avalia a influência. Geralmente, um diagnóstico de influência local é observado a partir de um gráfico de índices. O gráfico de índice da direção da curvatura, d_{\max} , correspondente ao máximo auto-valor de $\mathbf{F}_\theta = \Delta^\top \mathbf{h}_\theta^{-1} \Delta$, $C_{d_{\max}}$, avaliado em $\theta = \hat{\theta}$ e $\omega = \omega_0$, pode conter informação valiosa para diagnóstico. Bem como, o gráfico de índice de $C_i = 2|f_{ii}|$, na qual f_{ii} o i th elemento da diagonal de F_θ , pode ser utilizado como uma técnica de diagnóstico para avaliar a existência de informações influentes quando $\hat{C}_i > 2 \cdot \bar{C}$, na qual $\bar{C} = \sum_{i=1}^n \hat{C}_i / n$, são ditas potenciais influências.

No modelo ZARBS, temos que a forma de Δ_β , Δ_α e Δ_γ são dados por, respectivamente, $\Delta_\beta = \tilde{X}^T A_\kappa \begin{bmatrix} (y'_i - \mu'_i) \delta_{ii}^n \end{bmatrix}$, $\Delta_\alpha = \tilde{Z}^T B_\kappa \begin{bmatrix} (y'_i - \sigma'_i) \delta_{ii}^n \end{bmatrix}$ e $\Delta_\gamma = \tilde{W}' C \begin{bmatrix} (y'_i - v'_i) \delta_{ii}^n \end{bmatrix}$, avaliados em $\hat{\theta}$ e $\omega_0 = 1_n$, que é, um vetor coluna $n \times 1$ com todos os elementos iguais a 1 e δ_{ii}^n é uma matriz diagonal $n \times n$.

4.3 Exemplo de Aplicação

Como exemplo de aplicação usaremos o conjunto de dados que previamente já foi estudado pelo instituto de biomedicina da Universidade de São Paulo (USP), contendo 364 observações. O prévio trabalho tinha por objetivo estudar a produção de fumonisina por parte do fungo fusarium verticillioides mais frequentemente encontrado em grãos de milho, trigo e outros cereais. A fumonisina é o membro mais prevalente da família de toxina, (ROCHA *et al.*, 2017).

Das 364 observações do conjunto de dados, 317 delas apresentam resposta com valor atribuído igual a 0 (zero), o que representa o um total de 87%, enquanto que os demais valores da resposta compõe os 13% restantes. Estudaremos, neste contexto, a relação entre as variáveis Fusarium Verticillioides, (x_1), Tempo em dias, (x_2), Precipitação, (x_3). Atividade da água, (x_4) e a Produção da toxina Fumonisina (FB2) que é nossa resposta, (y).

O objetivo desta aplicação será um estudo comparativo de três modelos específicos para dados com alta quantidade de zeros. Dentre estes modelos utilizaremos a Normal Inversa Zero Ajustada, referenciada aqui como ZAIG, e distribuição Gama Zero Ajustada,

aqui chamada de ZAGA e a distribuição em questão, a distribuição Birnbaum-Saunders Reparametrizada Zero Ajustada, aqui chamada de ZARBS, (TOMAZELLA *et al.*, 2019).

Para todos os modelos que serão ajustados, observaremos medidas como AIC e SBC a fim de determinar qual dos modelos apresenta melhor ajuste. Também mostraremos os valores-p dos parâmetros estimados a fim de saber que variáveis contribuem no modelo.

De maneira a comparar outras performances, utilizaremos diferentes funções de ligação para o modelo ZARBS. As funções em questão serão: Logito, Probit e Complemento log-log, todas aplicadas ao parâmetro v . Além de usar a função log para μ e função identidade para o parâmetro de precisão σ .

O modelo ajustado então segue a relação funcional dada a por:

$$\begin{aligned} g_1(\mu_i) &= \eta_i = f_1(X_i; \beta), \\ g_2(\sigma_i) &= \tau_i = f_2(Z_i; \alpha), \\ g_3(v_i) &= \varepsilon_i = f_3(W_i; \gamma), \end{aligned} \tag{4.14}$$

que, no caso da ZARBS, é expressa por:

$$\begin{aligned} \log(\mu_i) &= \text{intercpto} + \beta_1 * F_vertc + \beta_2 * Dias_flor + \beta_3 * Precipt1 + \beta_4 * Aa, \\ \sigma_i &= \gamma, \\ \log\left(\frac{v_i}{1-v_i}\right) &= -\text{intercpto} + \alpha_1 * F_vertc + \alpha_2 * Dias_flor + \alpha_3 * Precipt1 + \alpha_4 * Aa, \end{aligned} \tag{4.15}$$

4.3.1 Resultados Abordagem Clássica

Tabela 1 – Estimativas dos Parâmetros para o modelo ZARBS.

Modelo	F. Ligação	Parâmetro	Variáveis	Estimativas	Erro Padrão	Valores-p
ZARBS	log	Média	intercepto	-8,509	15,928	0,593
			F_vertc	0,016	0,006	0,019
			Dias_flor	0,057	0,015	0,000
			Precipit1	-0,007	0,047	0,882
			Aa	3,900	16,048	0,808
ZARBS	logito	Proporção	intercepto	-7,938	14,660	0,588
			F_vertc	-0,025	0,007	0,000
			Dias_flor	-0,046	0,015	0,003
			Precipit1	0,174	0,047	0,000
			Aa	14,492	14,747	0,436
ZARBS	identidade	Precisão	-	1,132	0,234	0,000

Fonte: Elaborada pelo autor.

Tabela 2 – Estimativas dos Parâmetros para o modelo ZARBS.

Modelo	F. Ligação	Parâmetro	Variáveis	Estimativas	Erro Padrão	Valores-p
ZARBS	log	Média	intercepto	-8,509	15,928	0,593
			F_vertc	0,016	0,006	0,019
			Dias_flor	0,057	0,015	0,000
			Precipit1	-0,007	0,047	0,882
			Aa	3,900	16,048	0,808
ZARBS	probito	Proporção	intercepto	-3,068	8,292	0,711
			F_vertc	-0,141	0,003	0,000
			Dias_flor	-0,027	0,008	0,001
			Precipit1	0,098	0,024	0,000
			Aa	5,037	8,342	0,546
ZARBS	identidade	Precisão	-	1,132	0,234	0,000

Fonte: Elaborada pelo autor.

Tabela 3 – Estimativas dos Parâmetros para o modelo ZARBS.

Modelo	F. Ligação	Parâmetro	Variáveis	Estimativas	Erro Padrão	Valores-p
ZARBS	log	Média	intercepto	-8,509	15,928	0,593
			F_vertc	0,016	0,006	0,019
			Dias_flor	0,057	0,015	0,000
			Precipit1	-0,007	0,047	0,882
			Aa	3,900	16,048	0,808
ZARBS	cloglog	Proporção	intercepto	0,459	7,499	0,951
			F_vertc	-0,012	0,003	0,001
			Dias_flor	-0,027	0,008	0,000
			Precipit1	0,086	0,019	0,000
			Aa	0,909	7,544	0,904
ZARBS	identidade	Precisão	-	1,132	0,234	0,000

Fonte: Elaborada pelo autor.

Tabela 4 – Estimativas dos Parâmetros para o modelo ZAIG.

Modelo	F. Ligação	Parâmetro	Variáveis	Estimativas	Erro Padrão	Valores-p
ZAIG	log	Média	intercepto	-93,801	20,687	0,000
			F_vertc	0,042	0,011	0,000
			Dias_flor	0,120	0,022	0,000
			Precipit1	-0,135	0,064	0,034
			Aa	89,775	20,914	0,000
ZAIG	logito	Proporção	intercepto	-7,938	14,660	0,588
			F_vertc	-0,025	0,007	0,000
			Dias_flor	-0,046	0,015	0,003
			Precipit1	0,174	0,047	0,000
			Aa	14,492	14,747	0,436
ZAIG	identidade	Precisão	-	2,530	0,261	0,000

Fonte: Elaborada pelo autor.

A partir das Tabelas 1, 2, 3, 4, 5, podemos observar que nem todas as variáveis são significativas em todos os modelos (tome como referência um valor-p menor que 0.05 como significativo). Com elas, também podemos obter os coeficientes de cada uma das variáveis para os diferentes modelos que usam diferentes funções de ligação.

Além dessa informação também obtivemos o valor das estatísticas AIC e SBC para que possamos tomar a decisão de qual o melhor modelo. Lembrando que quanto

Tabela 5 – Estimativas dos Parâmetros para o modelo ZAGA.

Modelo	F. Ligação	Parâmetro	Variáveis	Estimativas	Erro Padrão	Valores-p
ZAGA	log	Média	intercepto	-30,138	18,500	0,004
			F_vertc	0,010	0,006	0,076
			Dias_flor	0,087	0,014	0,000
			Precipit1	0,034	0,041	0,409
			Aa	24,940	18,591	0,180
ZAGA	logito	Proporção	intercepto	-7,938	14,660	0,588
			F_vertc	-0,025	0,007	0,000
			Dias_flor	-0,046	0,015	0,003
			Precipit1	0,174	0,047	0,000
			Aa	14,492	14,747	0,436
ZAGA	identidade	Precisão	-	0,927	0,085	0,000

Fonte: Elaborada pelo autor.

menores estes valores, melhor será o modelo. E com isso, chegamos a conclusão de que o melhor modelo para os dados ajustados, sem análise de diagnóstico, seria o modelo ZAGA, pois é aquele que tem o menor AIC e SBC respectivamente, como mostrado na tabela 6.

Tabela 6 – Medidas AIC e SBC para os modelos ajustados

Modelo	AIC	SBC
ZARBS v logito	213.88	256.75
ZARBS v probito	212.35	255.32
ZARBS v cloglog	210.86	253.73
ZAIG	216.61	259.48
ZAGA	207.79	250.66

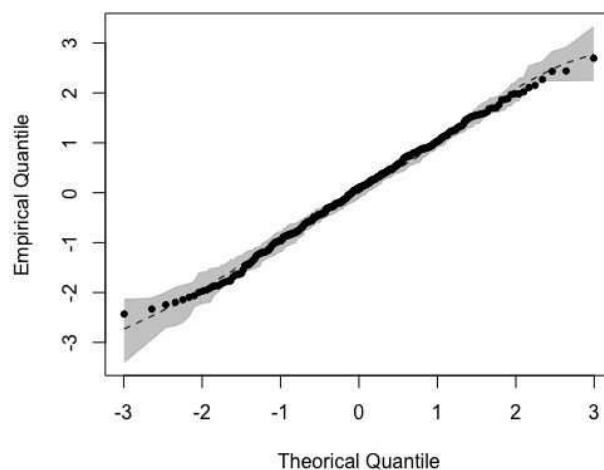
Fonte: Elaborada pelo autor.

4.3.2 Análise de Diagnóstico

Nos gráficos apresentados nas Figuras 4.3.1 e 4.3.2, correspondentes ao envelope simulado, para o modelo ZARBS, com confiança de 95% e ao gráfico de dispersão (resíduos vs índices das observações), conseguimos avaliar que a maior parte dos resíduos estão dentro dos intervalos do gráfico de envelope, satisfazendo a condição da

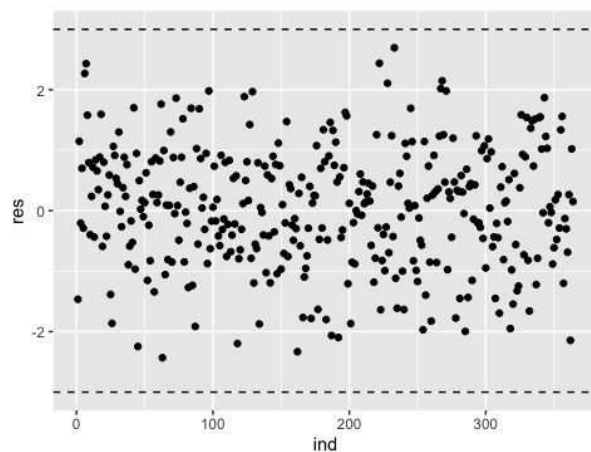
distribuição assumida, ainda temos que no gráfico de dispersão não observamos algum comportamento, diferente do aleatório, dos resíduos.

Figura 4.3.1 – Gráfico de Envelope Simulado.



Fonte: Tomazella *et al.* (2019).

Figura 4.3.2 – Gráfico de Dispersão dos Resíduos pelos Índices.

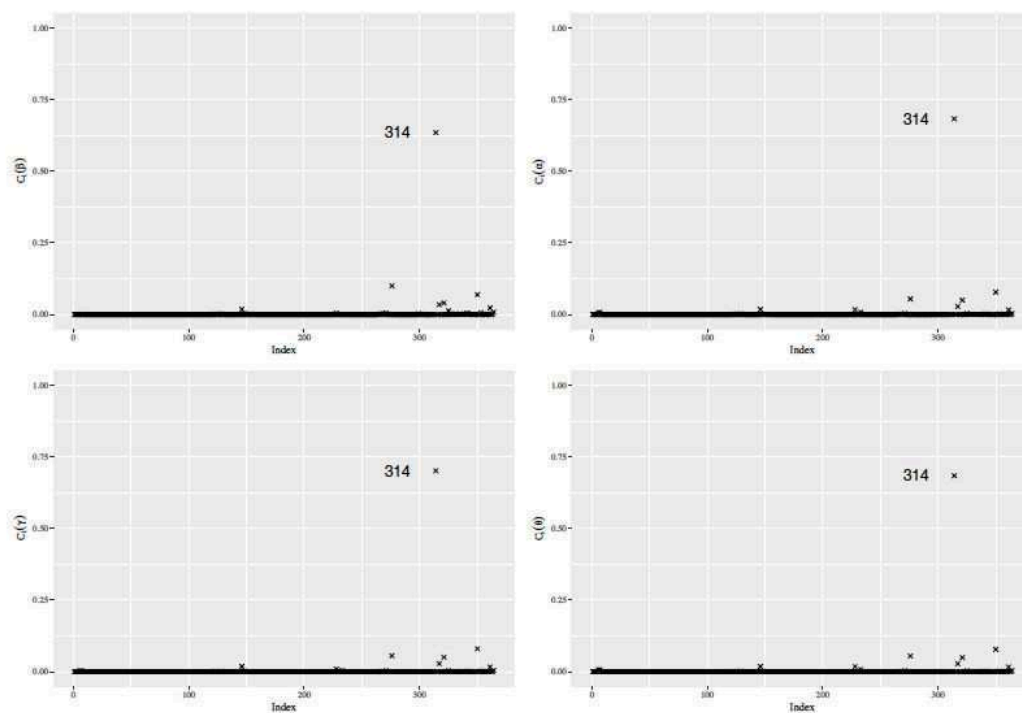


Fonte: Tomazella *et al.* (2019).

Observamos também se há possíveis pontos influentes no modelo e como resultado temos a Figura 4.3.3 que mostra que a observação 314 é um possível ponto influente neste modelo o que poderia ocasionar em alteração dos parâmetros estimados e suas

conclusões. A observação 314 possui quantidade de dias igual a 60, correspondente ao valor máximo presente no conjunto de dados, com precipitação igual a 5.57 (valor próximo ao primeiro quartil do conjunto de dados), atividade da água igual a 0.97 (valor mediano do conjunto de dados) e Fusarium Verticillioides de 45.5 (valor entre a mediana e o terceiro quartil) e o valor da resposta (a toxina Fumonisina (FB2)) igual a 0.02.

Figura 4.3.3 – Gráficos para Análise de Pontos Influentes.



Fonte: Tomazella *et al.* (2019).

Refizemos o modelo sem este possível ponto influente (observação 314) e observamos seu impacto nos parâmetros do modelo. Não observando impactos continuamos todas as observações, a partir dos resultados retiramos aquelas covariáveis não significa-

tivas e o modelo é apresentado a seguir:

$$\begin{aligned} \log(\mu_i) &= -30.975 + 0.017 * F_vertc + 0.075 * Dias_flor, \\ \sigma_i &= 1.457, \\ \log\left(\frac{v_i}{1 - v_i}\right) &= -9.928 - 0.026 * F_vertc - 0.044 * Dias_flor + 0.175 * Precipt1 \end{aligned} \tag{4.16}$$

UMA ABORDAGEM BAYESIANA PARA O MODELO DE REGRESSÃO ZARBS

Este capítulo tem por objetivo considerar uma abordagem Bayesiana para estimar os parâmetros do modelo considerado no Capítulo 3. A inferência bayesiana não é apenas uma alternativa cada vez mais popular à inferência freqüentista clássica, mas também é particularmente atraente para modelos hierárquicos ou multiníveis e para penalizar efeitos de regressão por meio de distribuições à priori adequadas. Muitas vezes é mais "tranquilo" implementar abordagem bayesiana, tendo em vista a álgebra envolvida. Além disso, abordagens totalmente bayesianas usando MCMC são atraentes em modelos de regressão flexíveis para obter intervalos de credibilidade a partir das amostras posteriores válidos para os estimadores em situações em que os intervalos de confiança falham para os estimadores de máxima verossimilhança correspondentes com base nas propriedades assintóticas. A análise bayesiana fornece um cenário conveniente para a estimativa de modelos de regressão aditiva generalizada complexa (GAMs). Como o poder computacional aumentou tremendamente na década passada, agora é possível resolver problemas inferenciais complicados, por exemplo, com a simulação de Monte Carlo em cadeias de Markov para praticamente qualquer computador.

5.1 O BAMLSS para o modelo de regressão ZARBS

Os termos de um preditor aditivo são mais comumente representados por abordagens de funções básicas. Isso leva a uma estrutura de modelo muito genérico e pode ser mais explorada porque cada termo pode ser transformado em uma representação de modelo misto (RUPPERT; WAND; CARROLL, 2003). Em um cenário totalmente bayesiano, essa generalidade permanece porque as prioris dos parâmetros também podem ser formalizados de maneira geral, por exemplo, atribuindo prioris normais aos coeficientes de regressão de termos suavizados ((FAHRMEIR *et al.*, 2013)). Neste trabalho vamos considerar o pacote BAMLSS para estimativa de modelos de regressão em uma estrutura bayesiana. Os parâmetros da distribuição podem capturar locação, escala, forma etc. e todos os parâmetros podem depender em termos aditivos complexos (fixos, aleatórios, suaves, espaciais etc.) semelhantes a um modelo aditivo generalizado na abordagem clássica (GAMLSS).

5.2 Distribuição a Posteriori

Para uma abordagem inferencial bayesiana completa, prioris precisam ser assumidas para os coeficientes dos modelos de regressão β_{jk} . Para ser o mais flexível possível faz-se uso de prioris gerais. Como uma priori geral entende-se aquela que possui grande variabilidade de forma que seja não informativa. Para o caso em estudo a informação a priori é dada por

$$P_{jk}(\cdot) = \pi_{jk}(\beta_{jk}, \tau_{jk}, \alpha_{jk}), \quad (5.1)$$

em que $\tau_{jk} = (\tau_1, \tau_2, \dots, \tau_k)$ é o vetor de todos os hiperparâmetros usados para as funções a priori $\pi_{jk}(\cdot)$, sabe-se ainda que τ_{jk} são tipicamente variâncias, por exemplo, que explicam o grau de suavização de $f_{jk}(\cdot)$ ou ainda o tamanho da correlação entre as covariáveis, similarmente, α é o conjunto de todos os valores fixos das funções a priori ($\alpha_\beta = m$, média a priori para β , e em muitas aplicações $m = 0$ em uma vaga especificação a priori e um parâmetro de precisão grande), $\beta_{ik} \sim N(0, \tau_{ik}^2 I)$ representa os efeitos aleatórios fixos iid.

A função de verossimilhança, conforme mostrada em (4.3), é repetida aqui de

forma a facilitar o entendimento desse passo para a construção da posteriori e sua forma é dada a seguir:

$$L(\mu, \sigma, \nu) = \prod_{i=1}^n f_y(y_i | \mu, \sigma, \nu) = \nu^{n_0} [1 - \nu]^{n - n_0} \prod_{i=1}^n f_T(y_i | \mu, \sigma)^{1 - I(y_i=0)}, \quad (5.2)$$

tendo em mãos a forma da verossimilhança, a distribuição a priori, podemos então chegar a distribuição a posteriori a partir do produto da priori e a verossimilhança. No entanto maximizar a posteriori ou a log-posteriori são equivalentes, portanto para encontrar as estimativas dos parâmetros basta maximizar a função de log posteriori apresentada abaixo:

$$\log \pi(\beta, \tau; y, X, \alpha) \propto l(\beta; y, X) + \sum_{k=1}^K \sum_{j=1}^{J_k} [\log p_{jk}(\beta_{jk}; \tau_{jk}, \alpha_{jk})]. \quad (5.3)$$

Assumindo distribuições a priori $\pi_{jk}(\cdot)$ para os componentes individuais produzem o log-posterior dada por:

$$\log(\pi(\beta, \mu, y, X, Z, \alpha, \gamma)) = l(\nu) + l(\sigma, \mu) + \sum_{y_i > 0} \log(\pi_{jk}(\beta_{jk}, \tau_{jk}, \alpha_{jk})).$$

O principal componente dos algoritmos do modelo de regressão é a função de densidade de probabilidade ou, por razões computacionais, seu logaritmo. A estimativa normalmente requer para avaliar o logaritmo da função de verossimilhança que foi dada no Capítulo 3 e será repetida aqui para facilidade do leitor:

$l(\theta) = l(\nu) + l(\sigma, \mu)$, em que:

$$\begin{aligned} l(\nu) &= n_0 \log(\nu) + [n - n_0] \log(1 - \nu) \\ l(\sigma, \mu) &= [n - n_0] c(\mu, \sigma) - \frac{3}{2} \sum_{y_i > 0} \log(y_i) - \frac{[\delta + 1]}{4\mu} \sum_{y_i > 0} y_i - \sum_{y_i > 0} \frac{\mu \sigma^2}{4[\sigma + 1]y_i} \\ &\quad + \sum_{y_i > 0} \log\left(y_i + \frac{\mu \sigma}{[\delta + 1]}\right), \end{aligned} \quad (5.4)$$

em que $c(\mu, \sigma) = -\frac{1}{2} \log(16\pi) + \frac{\sigma}{2} - \frac{1}{2} \log(\mu) + \frac{1}{2} \log(\sigma + 1)$. Agora considere que y_1, \dots, y_n , iid, seja uma sequência de variável aleatória tal que $Y_i \sim ZARBS(\mu_i, \sigma_i, \nu_i)$ e

suponha ainda que os parâmetros que representam a média, a variância e a mistura de Y_i satisfaçam as relações funcionais:

$$\begin{aligned} g_1(\mu_i) &= \eta_i = f_1(X_i; \beta), \\ g_2(\sigma_i) &= \tau_i = f_2(Z_i; \alpha), \\ g_3(v_i) &= \varepsilon_i = f_3(W_i; \gamma), \end{aligned} \tag{5.5}$$

para $i = 1, \dots, n$ na qual $\beta = [\beta_1, \dots, \beta_k]^T$, $\alpha = [\alpha_1, \dots, \alpha_q]^T$ e $\gamma = [\gamma_1, \dots, \gamma_r]^T$ que representam os vetores de parâmetros desconhecidos a serem estimados. Esta metodologia será aplicada no mesmo conjunto de dados do Capítulo 4 a fim de observar as principais diferenças entre a abordagem clássica e bayesiana.

5.3 Aplicação

Nesta seção, vamos utilizar os dados da Seção 3.3 considerando uma abordagem Bayesiana. Para isto, iremos usar o pacote `bamlss`. Primeiro, iremos considerar o modelo com todas as variáveis para verificarmos se os coeficientes selecionados serão os mesmos do ajuste com a abordagem frequentista. Usando o comando `bamlss()` nós geramos uma cadeia de tamanho 10 000, com as primeiras 5 000 amostras descartadas e consideramos os saltos como 10.

Antes de começar a aplicação, vale ressaltar a priori utilizada pelo pacote é: $\tilde{\beta}_{jk} \sim N(0, \tau_{jk}^2 I)$ iid são os efeitos aleatórios. Esta priori é não informativa de forma a acoplar melhor as informações inerentes aos dados.

Nas Figuras A.0.1-A.0.4, apresentamos os gráficos de convergência para as estimativas dos coeficientes, ou seja, a trajetória dos parâmetros estimados via algoritmo MCMC, e os gráficos das funções de autocorrelação. Podemos observar que o comportamento das cadeias são similares para todos os parâmetros.

Na Tabela 7, apresentamos os intervalos de credibilidade de 95% para cada estimativa, juntamente com as estimativas de Bayes para os parâmetros que foram obtidos considerando a moda das amostras geradas à posteriori. Com os resultados apresentados na Tabela 7, podemos notar que para a modelagem da média apenas a variável `F_vertc` é significativa. Já para a modelagem dos zeros, temos que as variáveis

Tabela 7 – Estimativas dos coeficientes obtidas usando o pacote bamls.

Coeficientes	Valores		Intervalo de Credibilidade (95%)	
	Estimativas	Erros-Padrão	Limite Inferior	Limite Superior
Média				
β_0	-8.533	23.294	-55.150	34.145
β_1	0.016	0.011	-0.009	0.040
β_2	0.058	0.021	0.028	0.102
β_3	-0.007	0.081	-0.110	0.175
β_4	3.929	23.812	-41.300	52.925
Proporção				
γ_0	-7.938	17.573	-39.384	13.716
γ_1	-0.026	0.007	-0.038	-0.010
γ_2	-0.047	0.016	-0.081	-0.019
γ_3	0.175	0.051	0.110	0.291
γ_4	11.492	17.749	-10.070	43.840
Precisão				
σ	0.875	0.262	0.458	1.292

Fonte: Elaborada pelo autor.

F_vertc, Dias_flor e Precipit1 foram significativas. Desta forma, nota-se uma diferença entre os modelos selecionados pela abordagem clássica e a Bayesiana.

Na Tabela 8, apresentamos os resultados considerando apenas a variáveis que foram significativas. Além disso, nas Figuras B.0.1-B.0.3 apresentamos os gráficos com o comportamento das cadeias. Nota-se que todos os gráficos apresentam um comportamento similar e que indicam a convergências das cadeias.

$$\begin{aligned}
 \log(\mu_i) &= -4.722 + 0.057 * Dias_flor, \\
 \sigma_i &= 1.126, \\
 \log\left(\frac{v_i}{1-v_i}\right) &= 3.471 - 0.026 * F_vertc - 0.053 * Dias_flor + 0.167 * Precipit1
 \end{aligned}
 \tag{5.6}$$

Tabela 8 – Estimativas dos coeficientes obtidas usando o pacote bamls.

Coeficientes	Valores		Intervalo de Credibilidade (95%)	
	Estimativas	Erro padrão	Limite Inferior	Limite Superior
β_0	-4.722	0.9378	-6.359	-3.343
β_2	0.057	0.014	0.059	0.089
γ_0	3.471	0.641	2.456	5.282
γ_1	-0.026	0.006	-0.045	-0.013
γ_2	-0.053	0.011	-0.086	-0.027
γ_3	0.167	0.036	0.092	0.267
σ	1.126	0.248	0.535	1.444

Fonte: Elaborada pelo autor.

A Tabela 9 mostra a comparação entre as duas abordagens propostas. Observamos que a variável F_Vert (com o respectivo β igual a β_2) é significativa no modelo clássico, mas não no modelo Bayesiano. Isso faz com que o intercepto tenha grande mudança na equação que modela a média. Também vemos grande diferença na equação que modela a proporção, principalmente no intercepto também. Já a estimativa para o σ embora haja diferença esta se apresenta apenas na segunda casa decimal, sendo a variância em torno de 1 em ambos os casos. Na abordagem bayesiana temos mais uma observação possivelmente influente, mas que ao analisar graficamente não desconsideramos sua influência tomando como base a medida KL, que será mostrada a seguir. De maneira geral ambos modelos podem ser utilizados (clássico e bayesiano) para o conjunto de dados em estudo.

Tabela 9 – Comparação das estimativas dos parâmetros na abordagem Clássica vs Bayesiana

Variáveis	Clássico		Bayesiano	
	$\log(\mu_i)$		$\log\left(\frac{v_i}{1-v_i}\right)$	
Intercepto	-30.975	-4.722	-9.928	3.471
F-vertc	0.017	-	-0.026	-0.026
Dias-flor	0.075	0.057	-0.044	-0.053
Precept1	-	-	0.175	0.167
σ Clássico			1.457	
σ Bayesiano			1.126	

Fonte: Elaborada pelo autor.

5.3.1 Diagnóstico Bayesiano

Uma maneira de avaliar a influência de uma observação em um determinado modelo é através do diagnóstico global de forma que o efeito da remoção deste observação é estudado. Esta análise é baseada na divergência de Kullback -Leibler(K-L). Seja $K(\pi, \pi_{(-i)})$ a divergência K-L entre π e $\pi_{(-i)}$, na qual π denota a distribuição a posteriori de θ para todos os dados, e $\pi_{(-i)}$ denota a distribuição a posteriori de θ sem a i -ésima observação. A medida é dada por:

$$K(\pi, \pi_{(-i)}) = \int \pi(\theta | D) \log \left\{ \frac{\pi(\theta | D)}{\pi(\theta | D_{(-i)})} \right\} \partial \theta, \quad (5.7)$$

em que $K(\pi, \pi_{(-i)})$ mede o efeito de retirar a i -ésima observação do conjunto de dados completo na distribuição a posteriori conjunta de θ . A calibração pode ser obtida resolvendo a equação:

$$K(\pi, \pi_{(-i)}) = K(B(0.5), B(p_i)) = \frac{-\log[4p_i(1-p_i)]}{2}, \quad (5.8)$$

em que $B(p)$ denota uma distribuição Benoulli com probabilidade de sucesso p . Após manipulações algébricas, podemos mostrar que $p_i = 0.5 \left[1 + \sqrt{1 - \exp[-2K(\pi, \pi_{(-i)})]} \right]$ implicando que $0.5 \leq p_i \leq 1$. Desta maneira então, se $p_i \gg 0.5$ então a i -ésima observação é influente. A equação a seguir mostra que esta medida pode ser expressa em termos da esperança a posteriori:

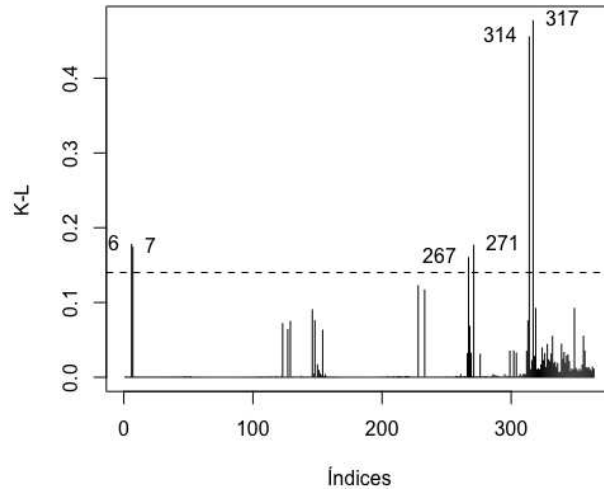
$$\begin{aligned} K(\boldsymbol{\pi}, \boldsymbol{\pi}_{-i}) &= \log E_{\boldsymbol{\theta}|D} [f(y_i | \boldsymbol{\theta})]^{-1} + E_{\boldsymbol{\theta}|D} [\log [f(y_i | \boldsymbol{\theta})]] \\ &= \log (CPO_i) + E_{\boldsymbol{\theta}|D} [\log [f(y_i | \boldsymbol{\theta})]], \end{aligned} \quad (5.9)$$

em que $E_{\boldsymbol{\theta}|D}(\cdot)$ denota a esperança com respeito à posteriori conjunta $\boldsymbol{\pi}(\boldsymbol{\theta} | D)$. Desta maneira, a medida K-L pode ser estimado através de reamostragens da distribuição a posteriori de $\boldsymbol{\pi}(\boldsymbol{\theta} | D)$ através do método MCMC. Portanto, se $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q$ é uma amostra de tamanho Q de $\boldsymbol{\pi}(\boldsymbol{\theta} | D)$, então:

$$K(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\pi}}_{-i}) = \log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log [f(y_i | \boldsymbol{\theta}_q)]. \quad (5.10)$$

A Figura 5.3.1 apresenta a medida Kullback-Leibler (K-L) usada para identificar pontos possivelmente influentes no modelo. Para esta medida consideramos que uma observação é influente quando seu valor supera 0.5.

Figura 5.3.1 – Medida KL para o modelo final ajustado.



Fonte: Elaborada pelo autor.

Neste caso, observamos que os valores 314 e 317 se aproximam de 0.5, no entanto não chegam a ultrapassar. Então consideramos que o modelo está bem ajustado e que não temos impactos de observações que poderiam alterar as estimativas dos parâmetros do modelo. Dessa forma, pontos como 314 e 317, de resíduo elevado e mais próximo

dos demais, mostram como o modelo proposto é robusto, isto é menos afetado por observações discrepantes.

CONCLUSÕES PRELIMINARES E PROPOSTA FUTURA

Neste trabalho, apresentamos a distribuição Birnbaum-Saunders, sua origem, suas principais características, bem como análise gráfica. No entanto, nos atemos principalmente ao modelo reparametrizado apresentado por (SANTOS-NETO *et al.*, 2012), quando apresentamos também a forma de reparametrização, suas características e análises gráficas fazendo uso do pacote RBS do software R, desenvolvido pelo mesmo autor do artigo supracitado. Seguimos para a abordagem considerando o modelo misto para a distribuição reparametrizada com zeros ajustados (ZARBS), tendo uma componente para modelar a proporção de zeros e outra modelando a resposta que seja diferente de zero. Para este último, trouxemos suas características, definimos o modelo de regressão e ferramentas de diagnóstico e, para fins de aplicação, trouxemos um exemplo fazendo uso desta distribuição considerando diferentes funções de ligação (logito, probito e complemento log-log e identidade para a variância). Além de comparar com outras duas distribuições frequentemente usadas para casos de zeros ajustados que são a Gama com Zeros Ajustados (ZAGA) e a Normal Inversa com Zeros Ajustados (ZAIG), apresentamos também seus resultados, que incluem análise de diagnóstico e equação final do modelo. Partimos, então, para a proposta deste trabalho que é apresentar uma abordagem bayesiana para este modelo. Vale ressaltar que esta parte do trabalho não se encontra na literatura, sendo algo inédito estudado. Para isso, fizemos uso do pacote BAMLSS,

que utiliza uma priori normal flat (uma priori não informativa) como padrão de entrada e fizemos um exemplo de aplicação agora Bayesiano para os mesmos dados, anteriormente trabalhados de maneira clássica. Também apresentando análises gráficas e equações finais. De maneira comparativa os resultados obtidos pelo modelo clássico e pelo modelo Bayesiano eram diferentes tendo diferentes variáveis, bem como os pesos das mesmas, que contribuem em ambos modelos.

De uma maneira geral, apresentamos mais uma distribuição a ser utilizada para casos onde temos alta concentração de zeros, cujos valores da resposta sejam não negativos e ressaltamos a versatilidade desta distribuição reparametrizada que comporta diferentes formas ao alterarmos seu parâmetro de forma, podendo ser mais ou menos simétrica o que colabora para uma maior versatilidade e aplicabilidade por conta disso. Além disso destacamos que o maior desafio da inferência Bayesiana é a disponibilização de ferramentas computacionais que são de fácil acesso e a incorporação do ZARBS ao BAMLSS contribui para que uma seleção de modelos estejam disponíveis ao usuário de inferência Bayesiana.

Considerando os resultados obtidos, há alguns problemas para serem discutidos em futuros trabalhos tais como outras técnicas de estudo de diagnóstico Bayesiano.

REFERÊNCIAS

ACHCAR, J. A. Inferences for the birnbaum—saunders fatigue life model using bayesian methods. **Computational statistics & data analysis**, Elsevier, v. 15, n. 4, p. 367–380, 1993. Citado na página 22.

AHMED, S. E.; BUDSABA, K.; LISAWADI, S.; VOLODIN, A. Parametric estimation for the birnbaum-saunders lifetime distribution based on a new parametrization. **Thailand Statistician**, v. 6, n. 2, p. 213–240, 2008. Citado na página 22.

ATKINSON, A. C. **Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis**. [S.l.]: Oxford, 1985. Citado na página 71.

BARROS, M.; PAULA, G. A.; LEIVA, V. An r implementation for generalized birnbaum-saunders distributions. **Computational Statistics & Data Analysis**, Elsevier, v. 53, n. 4, p. 1511–1528, 2009. Citado na página 22.

BAYES, T. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. **Philosophical transactions of the Royal Society of London**, The Royal Society London, n. 53, p. 370–418, 1763. Citado na página 39.

_____. Essay towards solving a problem in the doctrine of chances. **Biometrika**, v. 45, p. 293–315, 1958. Citado na página 33.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. **Journal of Applied probability**, Cambridge University Press, v. 6, n. 2, p. 319–327, 1969a. Citado nas páginas 21, 23, 27 e 28.

_____. Estimation for a family of life distributions with applications to fatigue. **Journal of Applied probability**, Cambridge University Press, v. 6, n. 2, p. 328–347, 1969b. Citado na página 22.

BROOKS, S.; SMITH, J.; VEHTARI, A.; PLUMMER, M.; STONE, M.; ROBERT, C. P.; TITTERINGTON, D.; NELDER, J.; ATKINSON, A.; DAWID, A. *et al.* Discussion on the paper by spiegelhalter, best, carlin and van der linde. **Journal of the Royal Statistical Society. Series B: Statistical Methodology**, Wiley-Blackwell, v. 64, n. 4, p. 616–639, 2002. Citado na página 44.

CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical Bayes methods for data analysis**. [S.l.]: Chapman & Hall/CRC, 2000. Citado na página 45.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M. D.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M.; GUO, J.; LI, P.; RIDDELL, A. Stan: A probabilistic programming language. **Journal of Statistical Software**, v. 76, p. 1, 2017. Citado na página 47.

CHEN, M.-H.; SHAO, Q.-M.; IBRAHIM, J. G. **Monte Carlo methods in Bayesian computation**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 46.

COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 48, n. 2, p. 133–155, 1986. Citado na página 71.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 30, n. 2, p. 248–265, 1968. Citado na página 71.

CRIBARI-NETO. **Engenharia de Avaliações com Base em Modelos GAMLSS**. Dissertação (Mestrado) — Programa de Pós-Graduação em Estatística, Universidade Federal de Pernambuco, Recife, 1 2010. Citado na página 50.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado na página 71.

DUPUIS, D. J.; MILLS, J. E. Robust estimation of the birnbaum-saunders distribution. **IEEE Transactions on Reliability**, IEEE, v. 47, n. 1, p. 88–95, 1998. Citado na página 22.

DÍAZ-GARCÍA, J.; LEIVA, V. A new family of life distributions based on elliptically contoured distributions. **J. Statist. Plan. Infer. J. Statist. Plan. Infer.**, v. 137, p. 1512–1513, 01 2005. Citado na página 22.

ENGELHARDT, M.; BAIN, L.; WRIGHT, F. Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. **Technometrics**, v. 23, p. 251–256, 1981. Citado na página 21.

FAHRMEIR, L.; KNEIB, T.; LANG, S.; MARX, B. **Regression**. Berlin Heidelberg: Springer-Verlag, 2013. Citado nas páginas 48 e 82.

GALEA, M.; LEIVA-SÁNCHEZ, V.; PAULA, G. Influence diagnostics in log-birnbaum-saunders regression models. **Journal of Applied Statistics**, Taylor & Francis, v. 31, n. 9, p. 1049–1064, 2004. Citado na página 22.

GELFAND, A. E.; DEY, D. K. Bayesian model choice: asymptotics and exact calculations. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 56, n. 3, p. 501–514, 1994. Citado na página 45.

GELFAND, A. E.; DEY, D. K.; CHANG, H. **Model determination using predictive distributions with implementation via sampling-based methods**. [S.l.], 1992. Citado na página 45.

HÄRDLE, W. **Nonparametric and Semiparametric Models**. Cambridge: Cambridge University Press, 1990. Citado na página 49.

KUNDU, D.; BALAKRISHNAN, N.; JAMALIZADEH, A. Bivariate birnbaum–saunders distribution and associated inference. **Journal of multivariate Analysis**, Elsevier, v. 101, n. 1, p. 113–125, 2010. Citado na página 22.

_____. Generalized multivariate birnbaum–saunders distributions and related inferential issues. **Journal of Multivariate Analysis**, Elsevier, v. 116, p. 230–244, 2013. Citado na página 22.

LEÃO, J.; LEIVA, V.; SAULO, H.; TOMAZELLA, V. *et al.* A survival model with birnbaum–saunders frailty for uncensored and censored cancer data. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 32, n. 4, p. 707–729, 2018. Citado na página 24.

LEIVA, V.; BARROS, M.; PAULA, G. A.; SANHUEZA, A. Generalized birnbaum–saunders distributions applied to air pollutant concentration. **Environmetrics: The official journal of the International Environmetrics Society**, Wiley Online Library, v. 19, n. 3, p. 235–249, 2008. Citado na página 22.

LEIVA, V.; RIQUELME, M.; BALAKRISHNAN, N.; SANHUEZA, A. Lifetime analysis based on the generalized birnbaum–saunders distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 52, n. 4, p. 2079–2097, 2008. Citado na página 22.

LEIVA, V.; SANHUEZA, A.; SEN, P. K.; PAULA, G. A. Random number generators for the generalized birnbaum–saunders distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 78, n. 11, p. 1105–1118, 2008. Citado na página 22.

LEIVA, V.; SANTOS-NETO, M.; CYSNEIROS, F. J. A.; BARROS, M. Birnbaum–saunders statistical modelling: a new approach. **Statistical Modelling**, Sage Publications Sage India: New Delhi, India, v. 14, n. 1, p. 21–48, 2014. Citado na página 23.

_____. Birnbaum–saunders statistical modelling: a new approach. **Statistical Modelling**, v. 14, n. 1, p. 21–48, 2014. Disponível em: <<https://doi.org/10.1177/1471082X13494532>>. Citado na página 31.

_____. A methodology for stochastic inventory models based on a zero-adjusted birnbaum-saunders distribution. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 32, n. 1, p. 74–89, 2016. Citado na página 24.

LIMA, L. P.; ANDRÉ, C. D. S.; SINGER, J. M. Modelos aditivos generalizados: metodologia e prática. **Revista Brasileira de Estatística**, v. 62, p. 37–69, 2001. Citado na página 49.

LU, M.-C.; CHANG, D. S. Bootstrap prediction intervals for the birnbaum-saunders distribution. **Microelectronics Reliability**, Elsevier, v. 37, n. 8, p. 1213–1216, 1997. Citado na página 22.

LUNN, D. J.; SPIEGELHALTER, D.; THOMAS, A.; BEST, N. The bugs project: Evolution, critique and future directions. **Statistics in Medicine**, v. 28, p. 3049, 2009. Citado nas páginas 44 e 47.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. Winbugs – a bayesian modelling framework: Concepts, structure, and extensibility. **Statistics and Computing**, v. 10, p. 325, 2000. Citado na página 47.

MANN, N.; SCHAFER, R.; SINGPURWALLA, N. **Methods for statistical analysis of reliability and life data**. Wiley, 1974. (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section). ISBN 9780471567370. Disponível em: <<https://books.google.com.br/books?id=yfdTAAAAMAAJ>>. Citado na página 21.

OWEN, W. J.; PADGETT, W. J. Accelerated test models for system strength based on birnbaum-saunders distributions. **Lifetime Data Analysis**, Springer, v. 5, n. 2, p. 133–147, 1999. Citado na página 22.

PLUMMER, M. **JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling**. 2003. Citado na página 47.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 24, 44 e 47.

RIECK, J. R.; NEDELMAN, J. R. A log-linear model for the birnbaum—saunders distribution. **Technometrics**, Taylor & Francis, v. 33, n. 1, p. 51–60, 1991. Citado na página 21.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape,(with discussion). **Applied Statistics**, v. 54, p. 507–554, 2005. Citado nas páginas 44, 47, 48, 49 e 51.

ROCHA, L. O.; REIS, G. M.; FONTES, L. C.; PIACENTINI, K. C.; BARROSO, V. M.; REIS, T. A.; PEREIRA, A. A.; CORRÊA, B. Association between fum expression and fumonisin contamination in maize from silking to harvest. **Crop protection**, Elsevier, v. 94, p. 77–82, 2017. Citado na página 72.

RUPPERT, D.; WAND, M.; CARROLL, R. J. **Semiparametric Regression**. Cambridge: Cambridge University Press, 2003. Citado nas páginas 48 e 82.

SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; AHMED, S. E. On new parameterizations of the birnbaum-saunders distribution. **Pakistan Journal of Statistics**, v. 28, n. 1, 2012. Citado nas páginas 11, 13, 23, 24, 25, 31 e 91.

SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; BARROS, M. On a reparameterized birnbaum-saunders distribution and its moments, estimation and applications. **REVSTAT-Statistical Journal**, v. 12, n. 3, p. 247–272, 2014. Citado na página 31.

SAULO, H.; LEÃO, J.; VILA, R.; LEIVA, V.; TOMAZELLA, V. On a bivariate birnbaum-saunders distribution parameterized by its means: features, reliability analysis and application. **arXiv preprint arXiv:1804.10820**, 2018. Citado na página 23.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the royal statistical society: Series b (statistical methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado na página 44.

TOMAZELLA, V.; PEREIRA, G. H.; NOBRE, J. S.; SANTOS-NETO, M. Zero-adjusted reparameterized birnbaum-saunders regression model. **Statistics & Probability Letters**, Elsevier, v. 149, p. 142–145, 2019. Citado nas páginas 23, 24, 65, 73, 77 e 78.

UMLAUF, N.; KLEIN, N.; SIMON, T.; ZEILEIS, A. **bamlss: A Lego Toolbox for Flexible Bayesian Regression (and Beyond)**. [S.l.], 2019. Disponível em: <<https://arxiv.org/abs/1909.11784>>. Citado na página 58.

UMLAUF, N.; KLEIN, N.; ZEILEIS, A. BAMLSS: Bayesian additive models for location, scale and shape (and beyond). **Journal of Computational and Graphical Statistics**, v. 27, n. 3, p. 612–627, 2018. Citado nas páginas 44, 48 e 51.

_____. Bamlss: Bayesian additive models for location, scale, and shape (and beyond). **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 27, n. 3, p. 612–627, 2018. Citado na página 53.

UMLAUF, N.; KNEIB, T.; KLEIN, N. **BayesX: R Utilities Accompanying the Software Package BayesX**. [S.l.], 2019. R package version 0.3-1. Disponível em: <<https://CRAN.R-project.org/package=BayesX>>. Citado nas páginas 44 e 47.

VILCA, F.; BALAKRISHNAN, N.; ZELLER, C. B. The bivariate sinh-elliptical distribution with applications to birnbaum–saunders distribution and associated regression and measurement error models. **Computational Statistics & Data Analysis**, Elsevier, v. 80, p. 1–16, 2014. Citado na página 23.

_____. A robust extension of the bivariate birnbaum–saunders distribution and associated inference. **Journal of Multivariate Analysis**, Elsevier, v. 124, p. 418–435, 2014. Citado na página 23.

WOOD, S. N. Thin-plate regression splines. **Journal of the Royal Statistical Society (B)**, v. 65, n. 1, p. 95–114, 2003. Citado nas páginas 47 e 51.

XIE, F.-C.; WEI, B.-C. Diagnostics analysis for log-birnbaum–saunders regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4692–4706, 2007. Citado na página 22.

YEE, T. W. The VGAM package for negative binomial regression. **Australian and New Zealand Journal of Statistics**, v. 61, 2020. Citado nas páginas 47 e 51.

GRÁFICOS DE CONVERGÊNCIA PARA O MODELO COM TODAS VARIÁVEIS

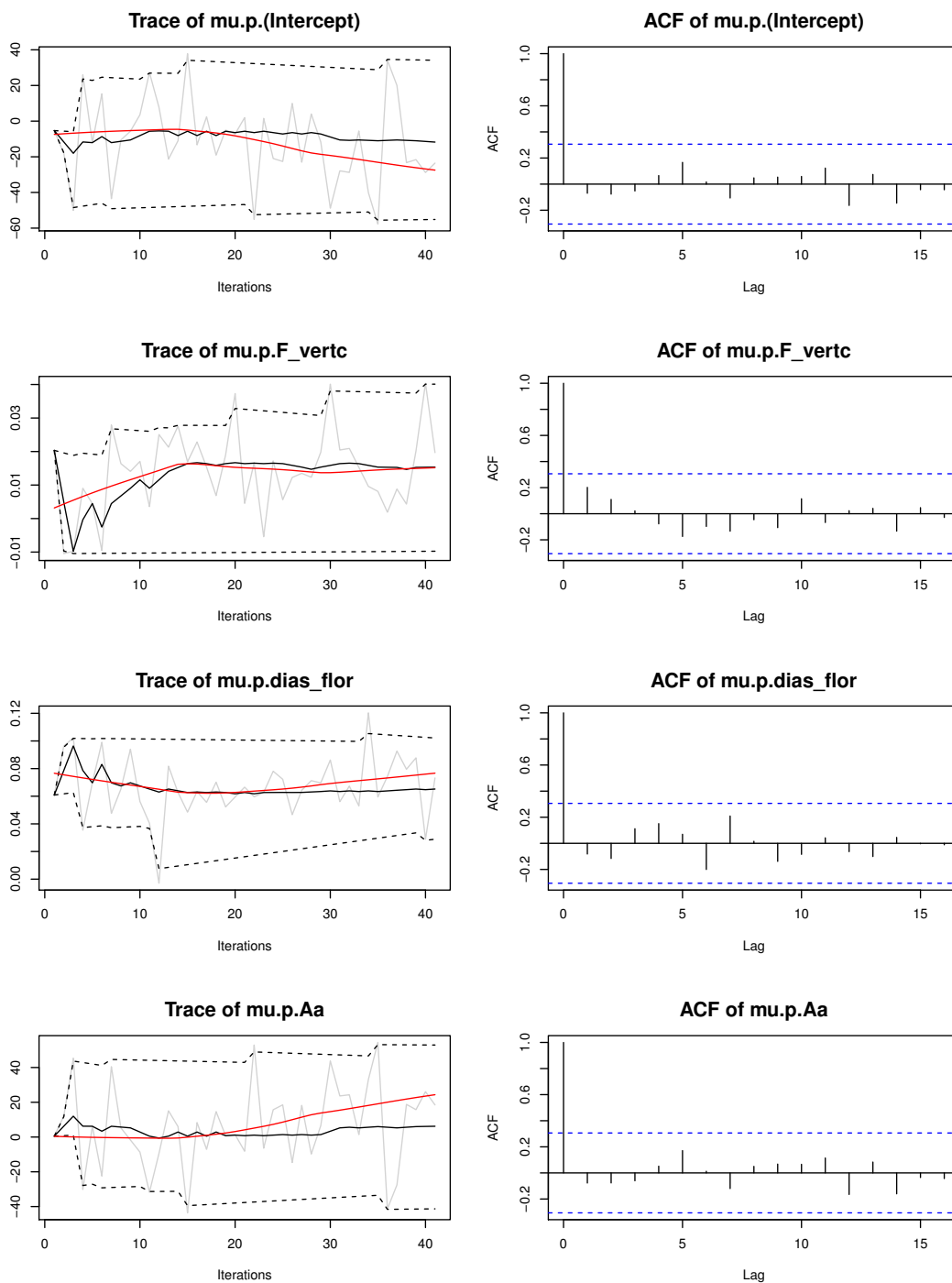


Figura A.0.1 – Gráficos para checagem da convergência para os parâmetros estimados.

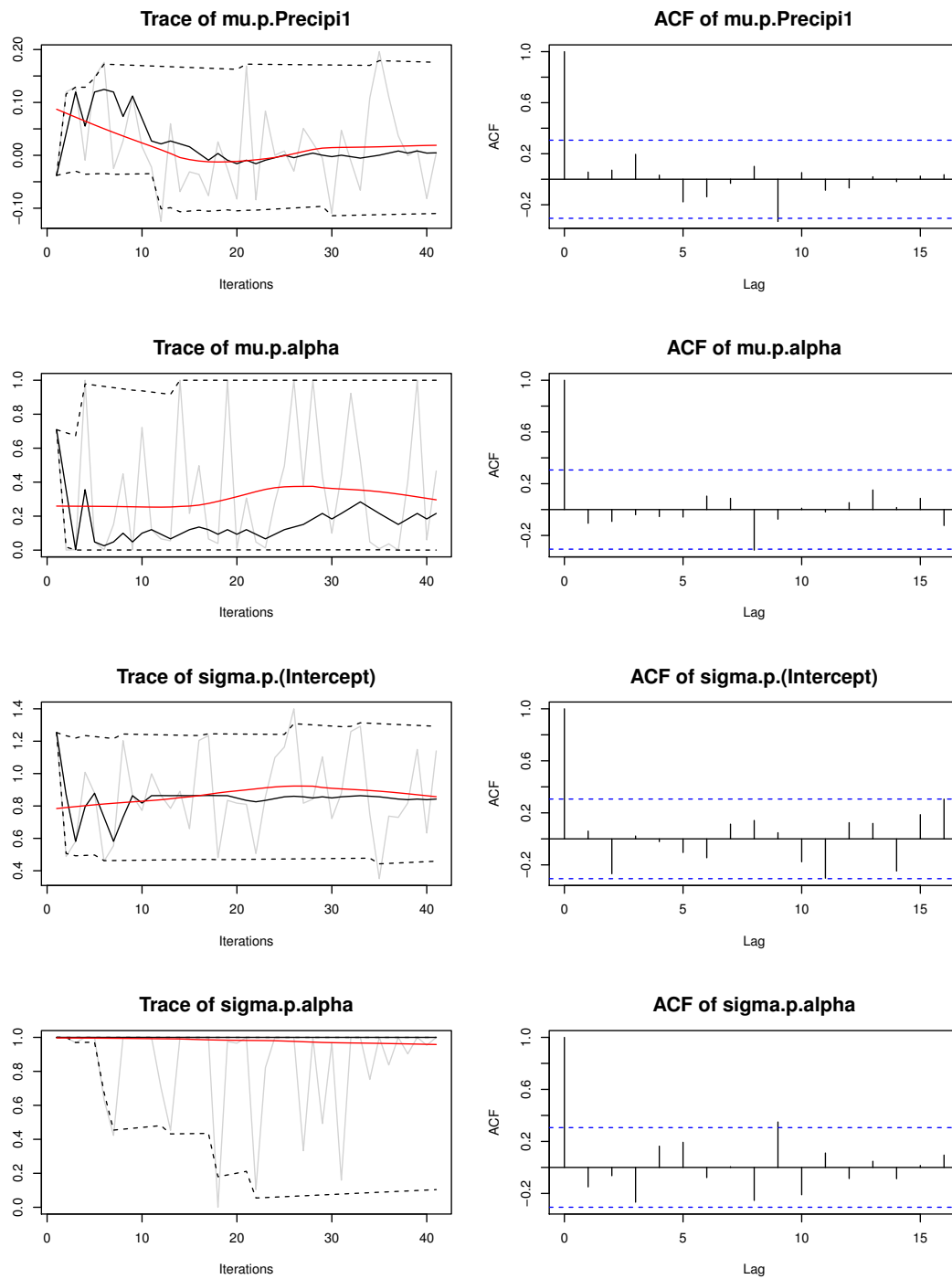


Figura A.0.2 – Gráficos para checagem da convergência para os parâmetros estimados.

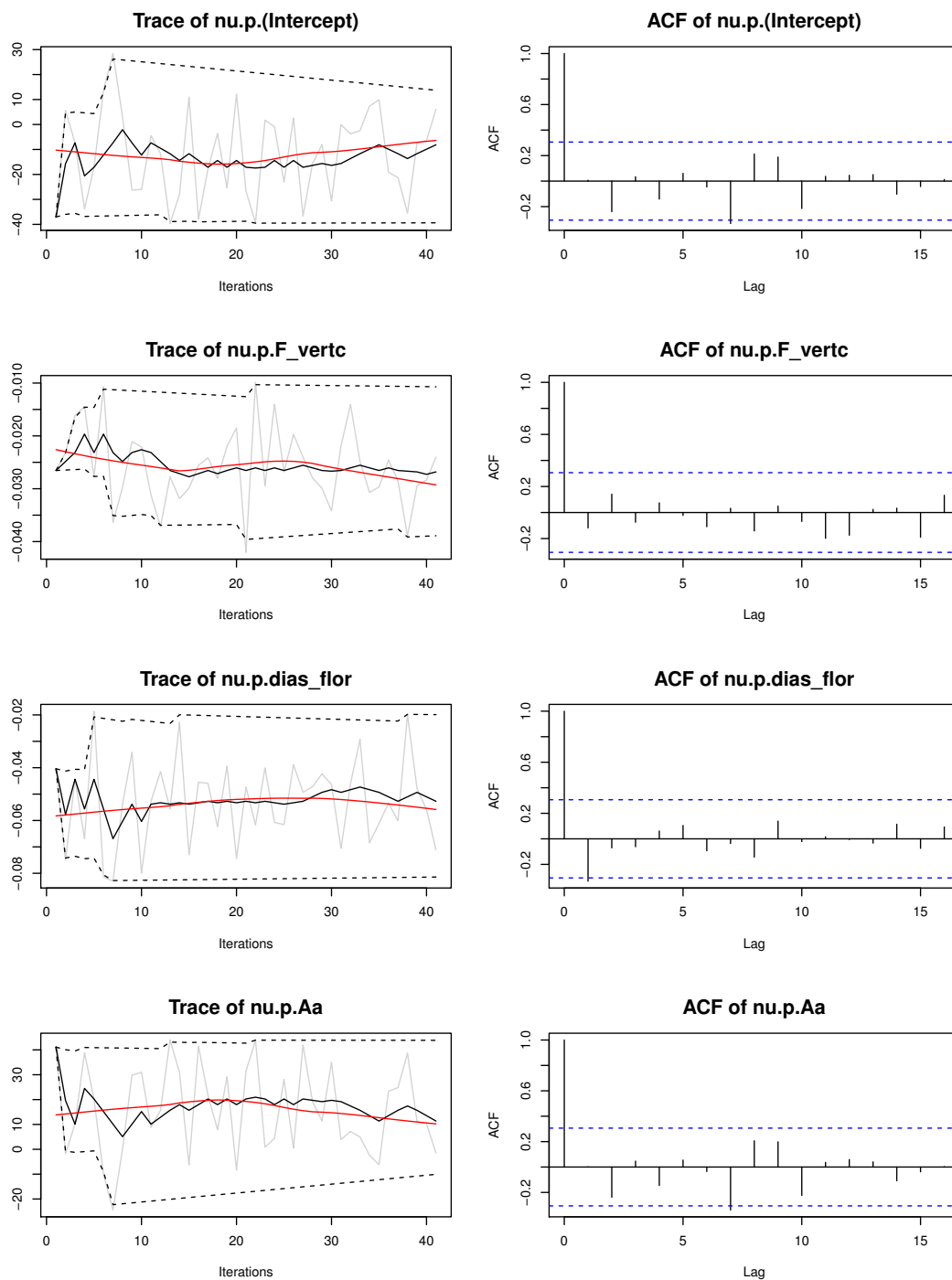


Figura A.0.3 – Gráficos para checagem da convergência para os parâmetros estimados.

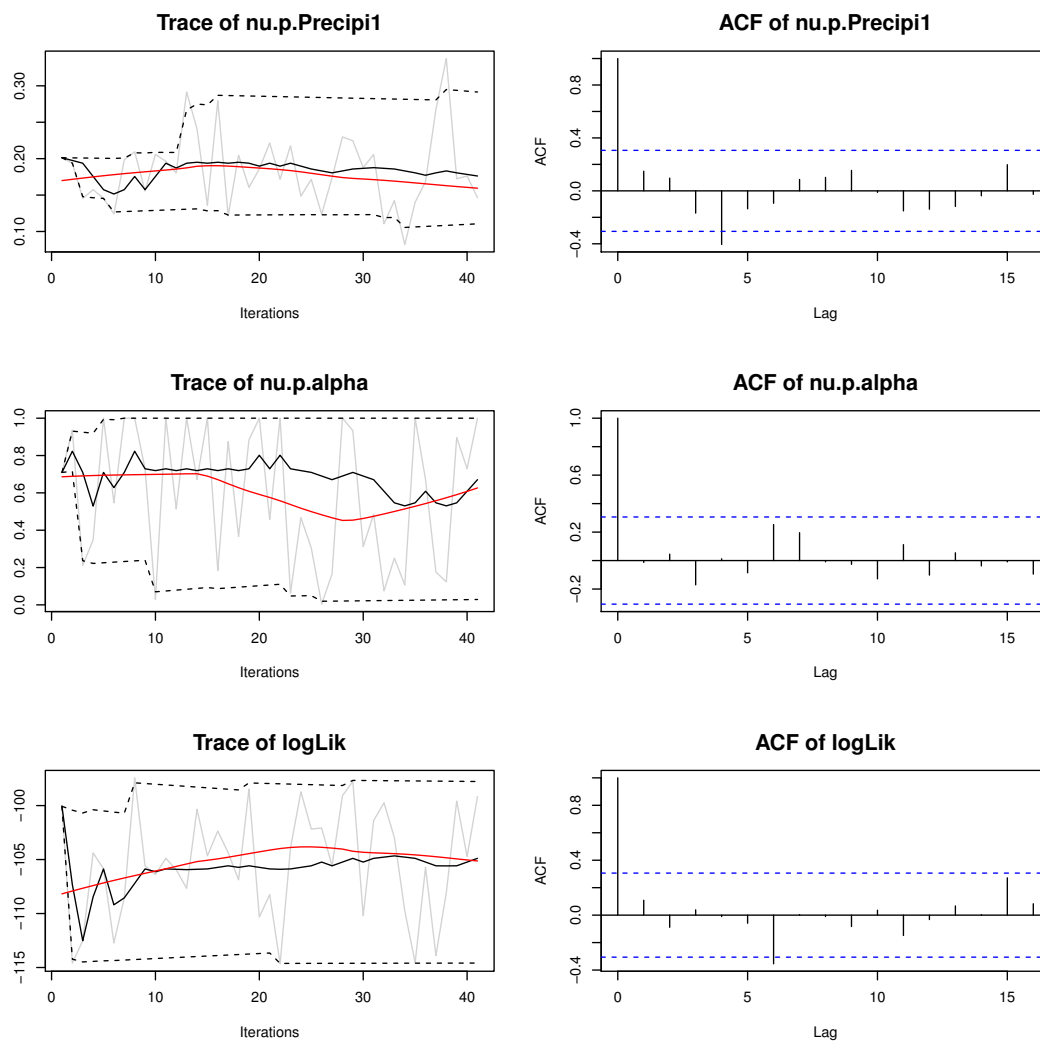


Figura A.0.4 – Gráficos para checagem da convergência para os parâmetros estimados.

APÊNDICE

B

GRÁFICOS DE CONVERGÊNCIA PARA O MODELO FINAL

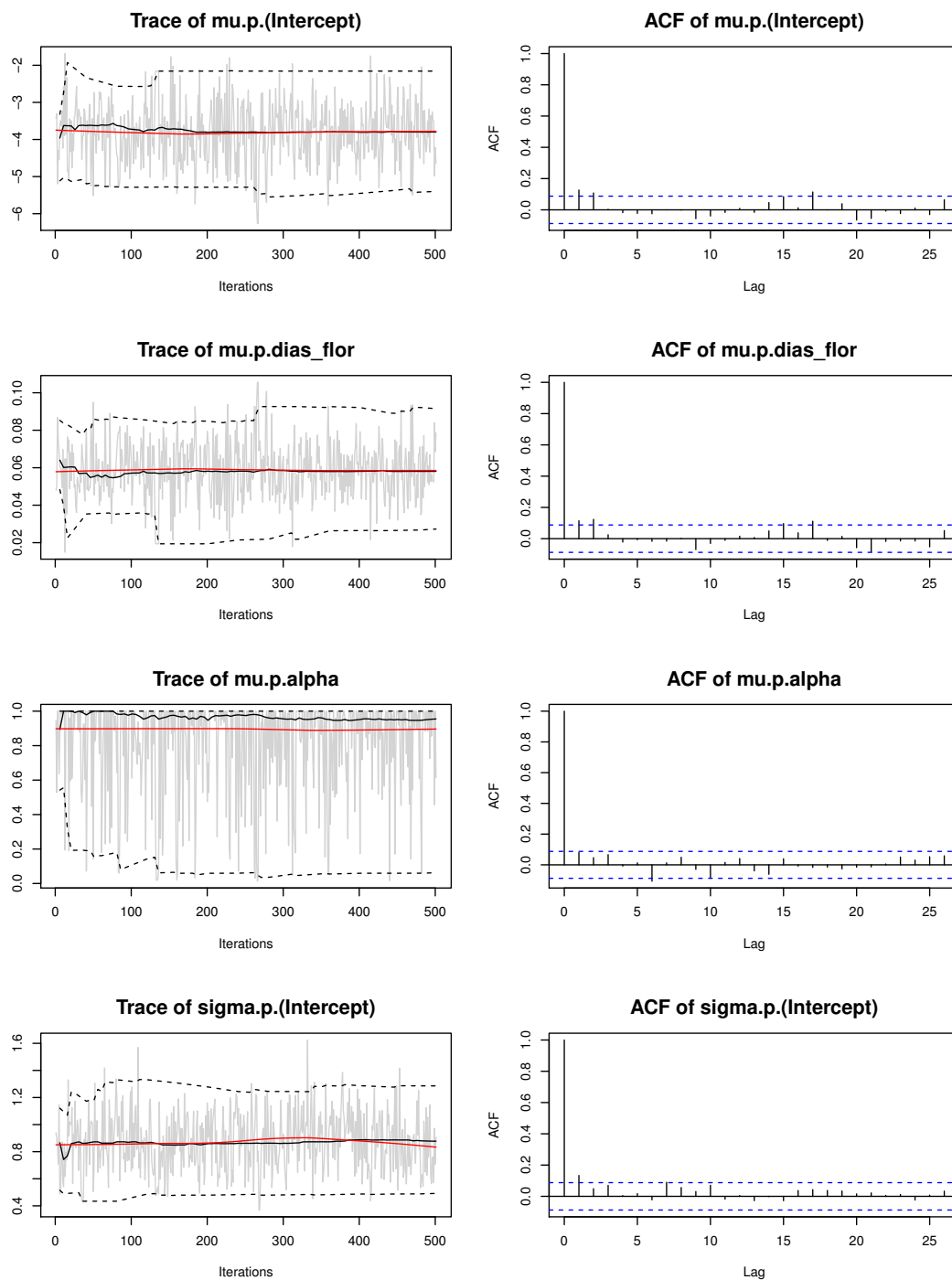


Figura B.0.1 – Gráficos para checagem da convergência para os parâmetros estimados.

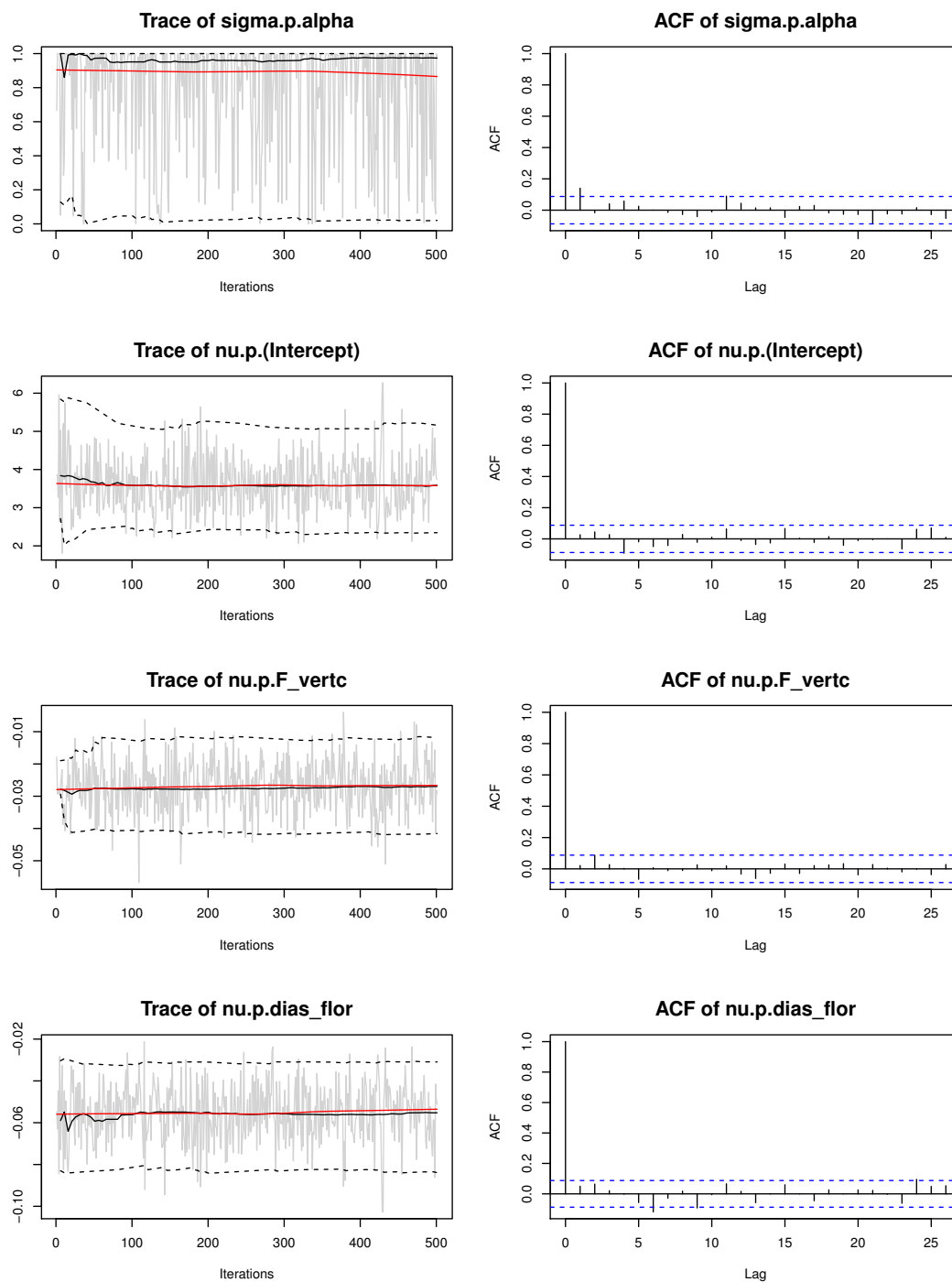


Figura B.0.2 – Gráficos para checagem da convergência para os parâmetros estimados.

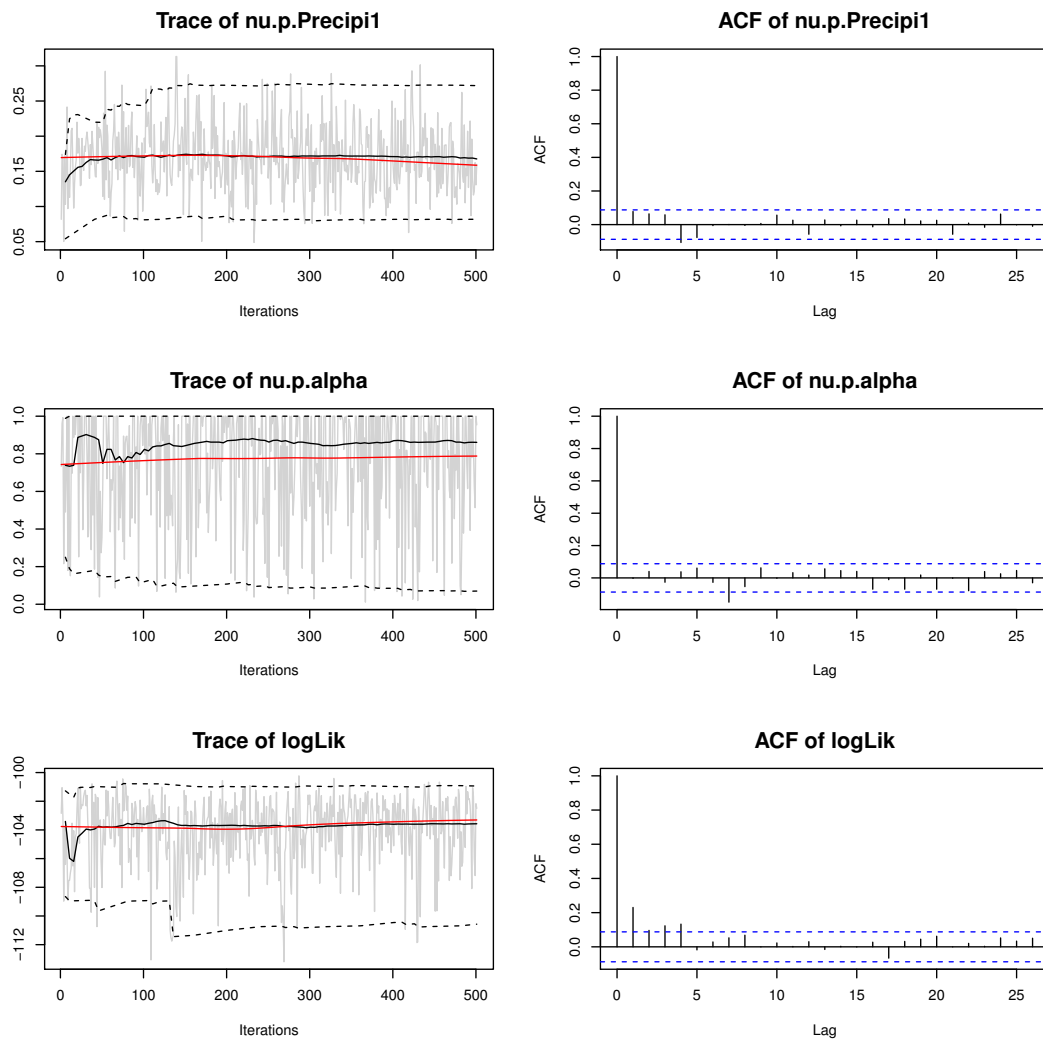


Figura B.0.3 – Gráficos para checagem da convergência para os parâmetros estimados.

