Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

Francisco Jucelino Matos Júnior

Modelo de Regressão sob Mistura de Escala Normal: Um enfoque não paramétrico para a Variável de Mistura

Recife
2017

Francisco Jucelino Matos Júnior

Modelo de Regressão sob Mistura de Escala Normal: Um enfoque não paramétrico para a Variável de Mistura

Dissertação apresentada ao programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Estatística. Área de Concentração: Estatística Aplicada.

ORIENTADOR: **Prof. Dr. Francisco José de Azevedo Cysneiros**
COORIENTADOR: **Prof. Dr. Aldo William Medina Garay**

Recife
2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

M433m Matos Júnior, Francisco Jucelino
Modelos de regressão sob mistura de escala normal: um enfoque não paramétrico para a variável de mistura / Francisco Jucelino Matos Júnior. – 2017.
53 f.: il., fig., tab.

Orientador: Francisco José de Azevedo Cysneiros.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências.

1. Análise de regressão. 2. Modelos de regressão. I. Cysneiros, Francisco José de Azevedo (orientador). II. Título.

519.536 CDD (23. ed.) UFPE- MEI 2017-76

FRANCISCO JUCELINO MATOS JÚNIOR

MODELOS DE REGRESSÃO SOB MISTURA DE ESCALA NORMAL: UM ENFOQUE NÃO PARAMÉTRICO PARA A VARIÁVEL DE MISTURA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 23 de fevereiro de 2017.

BANCA EXAMINADORA

Prof. Aldo William Medina Garay
UFPE

Prof.ª Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof. Roberto Ferreira Manghi
UFC

Dedico esse trabalho

Aos meus pais, Meire e Juscelino
À minha irmã Caroline e vó Marina
À toda a comunidade LGBTTQ
Aos meus amigos e familiares.

Agradecimentos

Queria agradecer a DEUS pelos momentos de consolo nessa nova jornada.

Aos meus pais, Meire e Juscelino, por todo amor, carinho e dedicação. Por todas as ligações de consolo e preocupação. Por me amarem e permitem ganhar asas e viver um dos melhores momentos da minha vida. Essa é mais uma conquista nossa.

À minha vó, Marina, por me apoiar e me lembrar que não é meu feitio desistir e que diploma nenhum te dá educação como cidadão e como ser humano, isso a gente aprende com a família e muitas vezes com a vida.

À minha irmã e cunhado, Carol e Ícaro, pelo apoio e amor incondicional.

À toda a minha família, sempre lembrando do orgulho que sentem por mim.

Aos meus colegas de turma, Olívia, Yuri, Rodney, Vinícius, Alejandro, Camila, Camila, Elisandra, Ramon, Jodavid e Walmir, por todo o companheirismo e auxílio. Ênfase para o melhor companheiro de apê Rodney, aos meus incríveis colombianos Olívia e Alejo, aos meus "mozões" Yuri e Vinícius por me mostrarem, mesmo quando eu já nem lembrava, a minha capacidade de ser maior.

À Lays e Jonas pelos melhores momentos da nossa vida de república, barzinhos e "rolês" noturnos.

Ao Emídio Lhewicheski, por compartilhar sua vida comigo e por querer saber ainda mais da minha. Por ser uma família pra mim em Recife, assim como eu pude ser um pouco da sua. Por dividir sonhos, tristezas e alegrias. Por cada conselho, por cada cerveja em Olinda e por todas as vezes que você me abraçou quando eu mais precisava de um carinho.

Aos amigos, Rubens e Bruna, pela recepção, amizade e carinho.

As minhas garotas de Fortaleza, Eduarda Quinderé, Ianna Queiroz, Ítalo Nogueira, Jackson Magalhães, Venícius Ferreira, Emanuel Romero e Victor Temóteo, pelo amor,

amizade, carinho, respeito e tudo que nossa amizade trouxe nesses dois anos em que estive distante. Amo vocês.

Aos meus orientadores, Prof. Dr. Aldo Medina e Prof. Dr. Francisco Cysneiros, pela dedicação nesse trabalho.

Aos meus professores do Programa de Pós-Graduação em Estatística da UFPE pelo os ensinamentos, em ênfase, ao Prof. Dr. Francisco Cribari pelo suporte e auxílio em vários momentos difíceis. És minha inspiração como pessoa e professor. Aos professores Prof. Dr. Alex Dias, Prof. Dr. Gauss Cordeiro e Prof. Dra. Audrey Cysneiros pela dedicação e tempo na minha formação durante as disciplinas cursadas.

Ao Prof. Dr. Juvêncio Nobre pelos ensinamentos e por ser um guia incrível nesse caminho que fiz até chegar esse momento. As palavras de bondade e amor nas minha dificuldades e a empatia demonstrada quando descobrimos um grande erro que cometemos, mas que ficará para trás e nos servirá de ensinamento.

Ao Prof. Dr. João Maurício Mota pelo amor incondicional pela minha pessoa e pela ajuda em todos esses dois anos, mesmo de tão distante. És uma das minha inspirações e peço desculpas pela distância estabelecida nos últimos sete meses. Muitas vezes, não queremos demonstrar fraqueza as pessoas que mais nos inspiram, porém sua solidariedade talvez fizesse um diferencial no meu caminho.

À Prof. Dra. Sílvia Maria de Freitas por ter me ensinado o caminho da pesquisa. Grande parte dessa dissertação foi construída com base no que aprendemos juntos no decorrer dos nossos dezoito meses juntos produzindo conhecimento. Espero um dia lhe devolver todo o carinho, amor e respeito que tivestes por mim. Meu maior exemplo de orientadora.

À todos os alunos do programa e amigos de Fortaleza, Recife, Olinda e da UFPE que de alguma forma ajudaram na minha chegada a este momento.

Ao CNPq, pelo apoio financeiro. Ao membros da banca examinadora, por suas valiosas sugestões.

"Quem gostou bate palma, quem não gostou paciência."

Tati Quebra Barraco

"If you can't love yourself, how in the hell you gonna love somebody else?"

Ru Paul

"It does not do to well on dreams and forget to live, remember that." (Albus Dumbledore em Harry Potter e a Pedra Filosofal)

J.K. Rowling

"The reason there is bigotry, sexism, racism, and homophobia is fear. People are afraid of their own feelings, afraid of the unknown and I am saying; don't be afraid."

Madonna

Resumo

Martin e Han (2016) propuseram o modelo de regressão linear (MRL-MEN), utilizando o algoritmo *Predicte Recursive* (PR) para estimar a distribuição da variável aleatória de mistura, considerando o parâmetro de escala conhecido e igual a um. Nesta dissertação estendemos o trabalho desenvolvido por Martin e Han (2016) propondo o modelo de regressão não linear (MRL-MEN) cujo erro tem distribuição de mistura de escala normal (MEN) não especificando uma distribuição para a variável de mistura. A principal motivação em trabalhar com a subclasse de distribuições MEN, é que esta permite trabalhar com distribuições com caudas mais pesadas, uma vez que os estimadores de máxima verossimilhança dos parâmetros do modelo são menos sensíveis a observações atípicas. Especificamente, desenvolvemos um processo para estimar os parâmetros no MRNL-MEN, considerando o parâmetro de escala conhecido e igual a um. Além disso, baseado em duas abordagens apresentadas em Efron (1979) e Louis (1982), estimamos a matriz de variâncias e covariâncias para os estimadores do modelo abordado. Por meio de estudos de simulação, avaliamos empiricamente as propriedades assintóticas dos estimadores em vários cenários, como por exemplo, as estimativas dos parâmetros na presença de observações atípicas e analisamos um conjunto de dados reais por meio da metodologia desenvolvida.

Palavras-chave: Mistura de Escala Normal. Modelo Não Linear. *Algoritmo Predictive Recursive*.

Abstract

Martin and Han (2016) proposed the linear regression model (LRM-SMN), using the Predictive Recursive (PR) algorithm to estimate the distribution of the mixing random variable, considering a scale parameter known and equal to one. In this present work, we extend the work developed by Martin and Han (2016) proposing the nonlinear regression model (NLRM-SMN) with a distribution error of a normal scale (SMN) not specifying a distribution for a mixture variable. The main motivation for working with the subclass of SMN, is that it allows practitioners to work with heavy tailed distributions, where maximum likelihood estimators of the model parameters are less sensitive to atypical observations. Specficaly, we developed a new estimation process to estimate the parameters in NLRM-SMN, considering known and equal to one. In addition, based on two approaches given in Efron (1979) and Louis (1982) we estimated the covariance matrix of the estimators of the model addressed. Through simulation studies, we evaluated empirically the asymptotic properties of the estimators in different scenarios, for example, the parameters estimates in the presence of outliers and analyzed a real data set through the developed methodology.

Keywords: Scale Mixture of Normal. Nonlinear Model. Predictive Recursive Algorithm.

Lista de Figuras

Lista de Tabelas

Sumário

CAPÍTULO 1

Introdução

1.1 Motivação

Situações em diversas áreas envolvem a análise das relações entre duas ou mais variáveis. A relação entre essas variáveis, denotadas por y e x, pode ser representada por um modelo linear. A coleção de ferramentas estatísticas que são usadas para modelar relações entre variáveis que estão relacionadas de maneira não determinística é chamada de análise de regressão ou modelos de regressão. Essa análise, utilizada para um grande número de aplicações, tem como objetivo relacionar uma variável resposta (y) a uma ou mais variáveis explicativas x. Porém, em alguns casos, essa relação pode não ser linear, nestes casos o modelo não linear se torna mais adequado, tendo em vista que vários fenômenos (sejam eles de qualquer natureza) e a relação entre as variáveis se dá em uma forma não linear. Modelos de regressão sob distribuição normal são amplamente difundidos na literatura. Em várias ocasiões, esses modelos não são adequados para alguns conjuntos de dados, como por exemplo, quando há a presença de observações atípicas.

Tendo em vista esse fato, pesquisadores da área vêm desenvolvendo metodologias mais sofisticadas para modelagem de dados. Como por exemplo, Andrews e Mallows (1974) que apresentaram uma sub-classe de distribuições denominadas de mistura de escala normal (MEN), que possui distribuições com caudas mais pesadas que a distribuição normal. Como casos particulares dessa sub-classe pode-se destacar as distribuições normal, Pearson tipo VII, t de Student, Slash, Normal contaminada, dentre outras. Outros pesquisadores da área também propuseram metodologias para modelagem dessa sub-classe e, desta forma, contornar o problema de modelagem em dados com presença de observações

atípicas. Pode-se destacar trabalhos como o de Meza et al. (2012), que desenvolveram métodos de estimação para modelos de regressão não lineares com efeitos mistos para algumas distribuições da subclasse MEN e Arellano-Valle (1994) que apresentou modelos para a classe de distribuições elípticas, a qual a sub-classe MEN está contida. Garay et al. (2015) desenvolveram o processo de estimação dos parâmetros via algoritmo EM (DEMPSTER et al., 1977) para modelos de regressão censurados considerando erros com distribuição sob a sub-classe MEN.

No contexto dos modelos de regressão linear para a sub-classe MEN, Martin e Han (2016) desenvolveram um processo de estimação para os parâmetros via algoritmo PR-EM (MARTIN; HAN, 2016) sem a especificação da distribuição da variável de mistura U e considerando o parâmetro de escala conhecido e igual a 1. Mais detalhes sobre o desenvolvimento do algoritmo PR-EM podem ser encontrados em Newton (2002), Ghosh e Tokdar (2006), Martin e Ghosh (2008), Tokdar et al. (2009) e Martin e Tokdar (2011).

O objetivo principal desse trabalho é generalizar o processo de estimação apresentado por Martin e Han (2016), para o caso dos modelos não lineares considerando parâmetro de escala conhecido e igual a 1.

Tem-se também os seguintes objetivos específicos:

1. Propor um estimador para o vetor de parâmetros $\boldsymbol{\beta}$ nos modelos de regressão não linear, cujos erros seguem uma distribuição de MEN sem supor uma distribuição em particular para a variável de mistura U;

2. Aproximar a matriz de variâncias e covariâncias desses estimadores, utilizando metodologias propostas na literatura;

3. Avaliar o comportamento assintótico desses estimadores e seu comportamento na presença de observações atípicas, via estudos de simulações;

1.2 Apresentação dos capítulos

A presente dissertação encontra-se dividida em quatro capítulos. Neste primeiro capítulo é apresentada a motivação deste trabalho e a definição dos objetivos do mesmo. No segundo capítulo defini-se a sub-classe MEN, tal como modelos de regressão linear e não linear nesta sub-classe de distribuições, sem a especificação da distribuição da variável de mistura e com parâmetro de escala conhecido, bem como o processo de estimação dos parâmetros e seus respectivos erros padrão assintóticos. No terceiro capítulo serão realizados estudos de simulação com objetivo de avaliar o comportamento dos estimadores discutidos no segundo capítulo, assim como a avaliação e comparação dos mesmos na presença de observações atípicas. Neste mesmo capítulo, as metodologias discutidas no segundo capítulo são aplicadas a um conjunto de dados reais. Por último, no quarto

capítulo são apresentadas as conclusões deste trabalho e propostas de trabalhos futuros a serem realizados.

CAPÍTULO 2

Modelo de regressão de mistura de escala normal via algoritmo PR-EM

Neste capítulo faz-se o desenvolvimento do processo de estimação dos parâmetros em modelos de regressão não linear (MRNL-MEN) sob distribuições de mistura de escala normal (MEN), sem assumir uma distribuição em particular para a variável de mistura. Inicialmente será apresentada uma pequena introdução a sub-classe de distribuições de MEN, assim como modelos de regressão linear (MRL-MEN) sob distribuições MEN discutidas em Martin e Han (2016). Em seguida, apresentam-se propostas de metologias que podem ser utilizadas para a estimação dos erros padrões dos estimadores do modelo.

2.1 Distribuição de Mistura de Escala Normal

Andrews e Mallows (1974) apresentaram uma subclasse de distribuições denominadas mistura de escala normal, que tem como casos particulares as distribuições normal, t de Student, Slash, Normal contaminada, Pearson tipo VII, dentre outras.

Uma variável aleatória (v.a.) Y tem distribuição MEN, com parâmetro de locação $\mu \in \mathbb{R}$ e parâmetro de escala σ^2 ($\sigma^2 > 0$) se sua função densidade de probabilidade (f.d.p.) é dada por:

$$f(y) = \int_0^\infty \phi\left(y; \mu, \kappa(u)\sigma^2\right) d\Psi(u; \boldsymbol{\nu}) \tag{2.1}$$

em que:

- $\phi(\cdot;\mu,\sigma^2)$ denota a f.d.p. de uma distribuição normal univariada com média μ e variância σ^2.
- $\kappa(\cdot)$ é uma função de ponderação positiva, por exemplo $\kappa(u) = 1/u$ que é considerada neste trabalho;
- U é uma v.a. positiva com função de distribuição acumulada (f.d.a.) e f.d.p., $\Psi(u,\boldsymbol{\nu})$ e $\psi(u,\boldsymbol{\nu})$, respectivamente, em que $\boldsymbol{\nu}$ é um escalar ou um vetor de parâmetros associados a distribuição da v.a. U.

Para a v.a. com densidade definida (2.1), denotada por $Y \sim MEN(\mu,\sigma^2,\Psi)$, tem a seguinte representação estocástica:

$$Y = \mu + \kappa^{1/2}(U)Z, \tag{2.2}$$

em que Z é uma variável aleatória com distribuição Normal com média zero e variância σ^2 e U é independente de Z, $U \perp Z$..

Basso et al. (2010) apresentaram algumas propriedades para as distribuições da sub-classe MEN, que são:

1. Se $E\left[\kappa^{1/2}(U)\right] < \infty$, então a esperança da v.a. Y é $E[Y] = \mu$;
2. Se $E[\kappa(U)] < \infty$, então a variância da v.a. Y é $V[Y] = E[\kappa(U)]\,\sigma^2$;
3. A função característica de Y é:

$$E\left[e^{itY}\right] = e^{it\mu}\int_0^\infty e^{-\frac{1}{2}\kappa(u)t^2\sigma^2}\psi(u)du.$$

Para mais detalhes veja Box e Tiao (1973), Andrews e Mallows (1974), Dempster et al. (1980), Fang et al. (1990) e Basso et al. (2010).

Dependendo da distribuição da variável de mistura U, pode-se obter uma distribuição em particular na sub-classe MEN. Assim, na Tabela 2.1 apresentamos algumas distribuições que fazem parte da sub-classe de MEN.

2.2 Caso 1: Modelo de regressão linear sob MEN

Considere o modelo de regressão linear,

Tabela 2.1: Distribuições de Mistura de Escala Normal univariadas.

Distribuição	Notação	$\psi(u)$
Normal	$N(\mu, \sigma^2)$	Degenerada em 1
t de Student	$t(\mu, \sigma^2, \nu)$	$Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ [a]
Slash	$SL(\mu, \sigma^2, \nu)$	$Beta\,(\nu, 1)$
Normal Contaminada	$NC(\mu, \sigma^2, \boldsymbol{\nu})$	Dicotômica [b]

[a] Considerando a distribuição $Gama(a, b)$ com média igual a $\frac{a}{b}$.

[b] Dicotômica com função de probabilidade (f.p.) $h(u; \boldsymbol{\nu}) = \xi \mathbb{I}_{\{\gamma\}} + (1 - \xi)\mathbb{I}_{\{1\}}$, $\boldsymbol{\nu} = (\xi, \gamma)$, $0 \leq \xi < 1$, $0 < \gamma \leq 1$, em que $\mathbb{I}_{\{B\}}(\cdot)$ é uma função indicadora do conjunto B.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.3}$$

em que $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ é o vetor de dimensão $n \times 1$ de respostas observadas, $\boldsymbol{X}$ é uma matriz $n \times p$ de posto completo, cujas linhas são denotadas por $\boldsymbol{x}_i$, $i = 1, \ldots, n$, contendo os valores de p variáveis explicativas, tal que $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]^\top$ e $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ é o vetor de erros aleatórios de dimensão $n \times 1$, tal que $\epsilon_i, i = 1, \ldots, n$ são independentes e identicamente distribuídos (i.i.d), seguindo distribuição MEN$(0, 1, \Psi)$ e neste trabalho considera-se a f.d.p. da variável de mistura U desconhecida.

Martin e Han (2016) propuseram um novo método de estimação dos parâmetros $\boldsymbol{\beta}$ em modelos de regressão linear sob distribuições de mistura de escala normal (MRL-MEN) com parâmetro de escala conhecido e igual a 1. A ideia principal do trabalho de Martin e Han (2016) é não assumir nenhuma distribuição para a variável de mistura U. Essa abordagem traz algumas vantagens para o modelo:

- **Redução do número de parâmetros a serem estimados**. Por não assumir uma distribuição em particular para a variável U, tal abordagem torna desnecessária a estimação do vetor de parâmetros $\boldsymbol{\nu}$.
- **Enfoque na caracterização da distribuição da variável resposta**. De forma geral, ao se trabalhar com MRL-MEN, há necessidade de assumir uma distribuição em particular para a variável de mistura U. Martin e Han (2016) utilizam o algoritmo *Predictive Recursive* (PR), proposto por Newton et al. (1998), para estimar a f.d.p. $\psi(u)$ da variável de mistura U. Este algoritmo, tem a vantagem de estimar $\psi(u)$ de forma não paramétrica e, consequentemente, a f.d.p. da variável resposta Y utilizando apenas os dados observados.

2.2.1 Algoritmo PR

O algoritmo PR foi proposto, inicialmente, como uma forma alternativa para o método *Markov chain Monte Carlo* (MCMC) na estimação da distribuição a posteriori em modelos de processo de mistura Dirichlet (NEWTON et al., 1998; NEWTON, 2002).

Este algoritmo tem grande utilidade no decorrer do desenvolvimento deste trabalho, pois possibilita estimar a f.d.p. da variável de mistura U e assim, utilizá-la para a construção de método de estimação do vetor de parâmetros $\boldsymbol{\beta}$.

Seja $(y_1, \ldots, y_n)$ amostra aleatória (a.a.) com f.d.p. $f(y)$ definida em (2.1) e $\psi(u)$ a densidade da variável U, ambas desconhecidas, com o componente $f(y|u) = \phi(y; \mu, \kappa(u))$ conhecido (no caso da sub-classe MEN, a densidade $f(y|u)$ possui distribuição normal). A seguir, apresentam-se os passos do algoritmo:

Passo 0: Escolhe uma f.d.p. inicial $\psi_0(u)$ e uma sequência de pesos w_i, tal que $0 < w_i < 1$, $i = 1, \ldots, n$, repetindo os seguintes passos para cada observação y_i;

Passo 1: Calcule a densidade

$$f_{i-1}(y_i) \equiv f_{\psi_{i-1}}(y_i) = \int_{\mathscr{U}} f(y_i|u)\psi_{i-1}(u)du.$$

Passo 2: Atualize a estimativa da densidade de mistura

$$\psi_i(u) = (1 - w_i)\psi_{i-1}(u) + w_i \frac{f(y_i|u)\psi_{i-1}(u)}{f_{i-1}(y_i)}.$$

Ao final, retorna $\psi_n(\cdot)$ e $f_{i-1}(y_i)$, as estimativas PR de $\psi(\cdot)$ e $f(y_i)$, $i = 1, \ldots, n$, respectivamente.

Segundo Martin e Han (2016),existem duas propriedades importantes, que devem ser destacadas, ao utilizar o algoritmo PR: a fácil implementação e agilidade computacional. Algumas propriedades de convergência acerca das estimativas obtidas através do algoritmo para grandes amostras são apresentadas em Tokdar et al. (2009).

Martin e Han (2016) descrevem algumas condições necessárias para a implementação do algoritmo PR.

- Os pesos w_i no algoritmo devem satisfazer duas condições, $\sum_{i=1}^{\infty} w_i = +\infty$ e $\sum_{i=1}^{\infty} w_i^2 < \infty$. Segundo Tokdar et al. (2009), a primeira condição é necessária para que $f_{n-1}(y_n)$ perca a influência de $f_0(y_1)$. Ao mesmo, os pesos w_i devem decair para zero para que $f_{i-1}(y_i)$ acumule informações a respeito da f.d.p. da variável resposta Y, a partir da amostra, em cada passo do algoritmo. Desta forma, a segunda condição garante uma taxa de decaimento dos pesos w_i.

- Para a inicialização do algoritmo, é necessário definir uma f.d.p. inicial $\psi_0(\cdot)$, tal que $\psi_0(u) > 0$, $\forall u \in \mathscr{U}$.

- As estimativas via algoritmo PR dependem da ordem como os valores amostrais são inseridos no algoritmo. Essa dependência pode ser contornada com várias permutações (aleatórias) da amostra, que consiste em reamostrar as observações y_i,

$i = 1, \ldots, n$ com reposição. Para cada amostra gerada, calcula-se as estimativas $f_{i-1}(y_i)$, $i = 1, \ldots, n$, referente a cada observação y_i, e $\psi_n(\cdot)$ e calcula-se a média de cada uma dessas estimativas. Martin e Tokdar (2012) afirmam que 25 permutações são suficientes e, dada a velocidade do algoritmo, esse procedimento de permutações não gera alto custo computacional.

2.2.2 Estimação do vetor de parâmetros $\boldsymbol{\beta}$

No contexto dos modelos de regressão, $f(y|u)$ depende de um vetor de parâmetros desconhecidos $\boldsymbol{\beta}$, isto é, $f(y|u) = f(y|\boldsymbol{\beta}, u)$, e o interesse está em obter estimativas para o vetor de parâmetro $\boldsymbol{\beta}$, Martin e Tokdar (2011) propuseram uma extensão do algoritmo PR que fornece uma aproximação da função de verossimilhança para $\boldsymbol{\beta}$ e permite, desta forma, fazer inferências sobre o mesmo. Seja $f_{i-1,\boldsymbol{\beta}}(y_i)$ a estimativa PR da f.d.p. dada em (2.1) baseada na amostra aleatória $(y_1, \ldots, y_n)$, logo a função de verossimilhança marginal PR de $\boldsymbol{\beta}$ é dada por:

$$L_{PR}(\boldsymbol{\beta}) = \prod_{i=1}^{n} f_{i-1,\boldsymbol{\beta}}(y_i).$$

Martin e Tokdar (2011) demonstraram que $L_{PR}(\boldsymbol{\beta})$ é uma aproximação para a função de verossimilhança marginal de $\boldsymbol{\beta}$.

Considere $f_{i-1,\boldsymbol{\beta}}(y_i)$, $i = 1, \ldots, n$, a estimativa obtida por meio do algoritmo PR para densidade $f(y_i)$ dado o vetor de parâmetros $\boldsymbol{\beta}$, baseada nos erros $\epsilon_1, \ldots, \epsilon_n$, em que $\epsilon_i = y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}$. A verossimilhança marginal PR de $\boldsymbol{\beta}$ pode ser reescrita como:

$$L_{PR}\left(\boldsymbol{\beta}\right) = \prod_{i=1}^{n} f_{i-1,\boldsymbol{\beta}}\left(y_i - \boldsymbol{x}_i^\intercal \boldsymbol{\beta}\right).$$

Martin e Han (2016) propuseram estimar o vetor de parâmetros $\boldsymbol{\beta}$ pela maximização da verossimilhança PR marginal conforme apresentado em (2.2.2), ou, equivalentemente, ao logaritmo da função de verossimilhança (log-verossimilhança) PR marginal, $\ell_{PR}\left(\boldsymbol{\beta}\right) = \log\left[L_{PR}\left(\boldsymbol{\beta}\right)\right]$. Para a implementação do algoritmo, utiliza-se as mesmas condições descritas na Seção .

Algoritmo híbrido PR-EM

O algoritmo *Expectation Maximization* (EM) foi introduzido por Dempster et al. (1977) e é utilizado para se obter o estimador de máxima verossimilhança de modelos estatísticos em situações em que a solução das equações de estimação não pode ser resolvida analítica ou diretamente.

Frequentemente é utilizado em problemas noas quais há dados faltantes ou em casos em que o modelo estatístico em estudo depende de variáveis latentes não observadas. O algoritmo consiste em considerar uma nova representação do modelo em "dados aumentados", que constitui em uma representação estocástica em termos de distribuições mais tratáveis, que dependem, em geral, de quantidades não observáveis, chamada de dados faltantes. A ideia do procedimento é obter uma log-verossimilhança aumentada, ou log-verossimilhança completa, supondo que as quantidades faltantes foram de fato observadas.

Sejam $\boldsymbol{y}_{obs}$ e $\boldsymbol{y}_{fal}$, dados observados e faltantes, respectivamente. Denota-se como $\boldsymbol{y}_c = \left(\boldsymbol{y}_{obs}, \boldsymbol{y}_{fal}\right)$ os dados completos. Desta forma, a função de log-verossimilhança dos dados completos é dada por:

$$\ell_c\left(\boldsymbol{\theta}|\boldsymbol{y}_c\right),$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros que se tem o interesse em estimar. Seja $\boldsymbol{\theta}^{(0)}$ o valores inicial de $\boldsymbol{\theta}$. O algoritmo divide-se em dois passos:

- **[Passo E:]** Cálculo da esperança condicional da log-verossimilhança dos dados completos condicionado aos dados observados $\boldsymbol{y}_o$ e a estimativa $\boldsymbol{\theta}^{(t)}$, dada por:

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = E\left[\ell_c\left(\boldsymbol{\theta}|\boldsymbol{y}_c\right)|\boldsymbol{y}_{obs}, \boldsymbol{\theta}^{(t)}\right].$$

- **[Passo M:]** Maximização de $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$ em função de $\boldsymbol{\theta}$.

Os passos E e M devem ser repetidos até que uma regra de convergência como, por exemplo, $||\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t+1)}||$ seja suficientemente pequena.

O algoritmo híbrido PR-EM, proposto por Martin e Han (2016) para a maximização da log-verossimilhança $\ell_{PR}(\boldsymbol{\beta})$. Martin e Han (2016) utilizaram o algoritmo EM com a finalidade de maximizar a $\ell_{PR}(\boldsymbol{\beta})$, obtida através do algoritmo PR. Reescrevendo $L_{PR}(\boldsymbol{\beta})$, tem-se

$$L_{PR}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \phi\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}; 0, u^{-1}\right) \times \frac{f_{i-1,\boldsymbol{\beta}},\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)}{\phi\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}; 0, u^{-1}\right)}.$$

Assim, a log-verossimilhança aproximada por meio do algoritmo PR é dada por:

$$\begin{aligned}\ell_{PR}(\boldsymbol{\beta}) = &\sum_{i=1}^{n} \log\left[\phi\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}; 0, u^{-1}\right)\right] \\ &- \sum_{i=1}^{n} \log\left[\frac{\phi\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}; 0, u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}}, y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}}\right].\end{aligned} \tag{2.4}$$

Martin e Han (2016) propuseram, para cálculo do passo E do algoritmo EM da função de log-verossimilhança em (2.4), a densidade dada por:

$$\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u) = \frac{\psi_{i-1,\boldsymbol{\beta}^{(t)}}(u)\phi\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}^{(t)}; 0, u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}^{(t)}\right)}, \tag{2.5}$$

em que $\boldsymbol{\beta}^{(t)}$ é uma atualização da estimativa obtida pelo algoritmo no passo t. Tal densidade pode ser vista como uma distribuição a posteriori para $\psi(\cdot)$. Desta forma, para o Passo E do algoritmo EM, deve-se obter a função $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$, definida por:

$$Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) + Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right).$$

em que

$$Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = \sum_{i=1}^{n}\int_{\mathscr{U}} \log\left[\phi\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}; 0, u^{-1}\right)\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du$$

e

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\sum_{i=1}^{n}\int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}},\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}\right)}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du.$$

Reescrevendo $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, tem-se

$$Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\frac{1}{2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}\right)^2\widehat{\omega}_i + c\left(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}^{(t)}\right)$$

em que

$$\widehat{\omega}_i = \int_{\mathscr{U}} u\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du. \tag{2.6}$$

Desta forma, $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ depende apenas de $\psi_{i-1,\boldsymbol{\beta}^{(t)}}$, $y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}$, das estimativas de $\boldsymbol{\beta}$ e da função $c\left(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}^{(t)}\right)$ não depende do vetor de parâmetros $\boldsymbol{\beta}$.

O objetivo ao considerar a função $Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ como uma soma de duas funções $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ e $Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, é desenvolver um processo de maximização apenas em $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, uma vez que,

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) \xrightarrow{\boldsymbol{\beta}^{(t)}\to\boldsymbol{\beta}} 0.$$

Esse resultado pode ser encontrado em Martin e Han (2016). Além disso, ao se escolher a densidade $\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)$, apresentada em (2.5), para obter a função $Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, torna o processo de maximização de $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, para a estimação de $\boldsymbol{\beta}$, em uma sequência de soluções de mínimos quadrados do modelo regressão linear ponderado pela matriz de pesos $\widehat{\boldsymbol{\Omega}}$, que é uma matriz bloco diagonal formada pelos pesos $\widehat{\omega}_1, \ldots, \widehat{\omega}_n$ dados em (2.6).

Algoritmo PR-EM

Dado $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$ e $\mathscr{U}, \psi_0, w_1, \ldots, w_n$, referente ao algoritmo PR, tem-se

Passo E: Calcular os pesos $\hat{\omega}_1, \ldots, \hat{\omega}_n$ utilizando o algoritmo PR com os erros $y_1 - \boldsymbol{x}_1^\top \boldsymbol{\beta}^{(t)}$, $\ldots$, $y_n - \boldsymbol{x}_n^\top \boldsymbol{\beta}^{(t)}$.

Passo M: Atualizar $\boldsymbol{\beta}^{(t)}$ pela maximização de $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ em $\boldsymbol{\beta}$, o que leva a expressão

$$\boldsymbol{\beta}^{(t+1)} = \left(\boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}^(}t)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{\Omega}^(}t)\boldsymbol{y},$$

em que $\widehat{\boldsymbol{\Omega}}^(t)$ é uma matriz diagonal formada pelos pesos $\hat{\omega}_1, \ldots, \hat{\omega}_n$ dado em (2.6) na t-ésima iteração.

Os passos E e M devem ser repetidos até a distância entre duas avaliações sucessivas de $\boldsymbol{\beta}$, seja menor que um valor suficientemente pequeno δ, ou seja, até que

$$||\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}|| < \delta.$$

2.3 Caso 2: Modelo de regressão não linear sob MEN

Considere o modelo de regressão não linear dado por:

$$y_i = g\left(\boldsymbol{x}_i; \boldsymbol{\beta}\right) + \epsilon_i, \ i = 1, \ldots, n \tag{2.7}$$

em que:

- $\mu_i = g\left(\boldsymbol{x}_i; \boldsymbol{\beta}\right)$ é uma função não linear contínua e diferenciável em $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ tal que a matriz de derivadas $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}$ tenha posto p $(p < n)$, para todo $\boldsymbol{\beta}$;
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ é o vetor de médias;
- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ é o vetor de respostas observadas;
- $\boldsymbol{x}_i$ tem os valores de p variáveis explicativas, tal que $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]^\top$;
- $\epsilon_i, i = 1, \ldots, n$ são os erros aleatórios i.i.d, seguindo distribuição MEN$(0, 1, \Psi)$ e a f.d.p. da variável de mistura U é considerada desconhecida;

Desta forma, a função de verossimilhança dos dados é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \int_0^\infty \phi\left(y_i - g\left(\boldsymbol{x}_i; \boldsymbol{\beta}\right), 0, u^{-1}\right) \psi(u) du.$$

Torna-se necessário, portanto estimar o vetor de parâmetros $\boldsymbol{\beta}$ e a f.d.p. da variável latente U, $\psi(u)$.

O objetivo desta seção é obter estimativas para o vetor de parâmetros $\boldsymbol{\beta}$ sem assumir uma distribuição para a variável de mistura U em MRNL-MEN. Tal abordagem se mostra útil em casos em que não há conhecimento suficiente sobre a verdadeira distribuição da variável resposta Y.

Partindo do pressuposto de que a f.d.p. da variável U é desconhecida, pode-se fazer o uso de ferramentas para obter, a partir dos dados amostrais, uma distribuição que melhor caracterize tais dados.

Há casos em que a verdadeira distribuição de U é desconhecida. Em tais situações, uma forma usual de determinar um bom modelo para os dados é especificar alguns modelos a serem ajustados e, dentre eles, escolher o que melhor se adeque aos dados utilizando critérios previamente definidos.

Uma das propostas desta dissertação é utilizar os dados amostrais para obter informações que caracterizem a distribuição da variável de mistura U e, consequentemente, a distribuição da variável resposta Y, em modelos de regressão não linear. Assim, partindo da ausência de informações sobre a densidade da variável U, busca-se apresentar uma forma de caracterizar a f.d.p. da variável resposta Y por meio de uma estimativa para a densidade da variável U baseada nos próprios dados amostrais.

2.3.1 Estimação do vetor de parâmetros $\boldsymbol{\beta}$

Seja o modelo de regressão dado em (2.7), em que a densidade de mistura $\psi(u)$ é desconhecida. Escrevendo o modelo em função dos erros, $\epsilon_i = y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta})$, $i = 1,\ldots,n$, tem-se

$$f(\epsilon_i) = \int_{\mathscr{U}} \phi\left(y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta});0,u^{-1}\right)\psi(u)du,\ i = 1,\ldots,n.$$

Desta forma, pode-se aplicar o algoritmo PR afim de estimar a f.d.p. $\psi(u)$. Seja $f_{i-1,\boldsymbol{\beta}}$ a estimativa do algoritmo PR para densidade $f(y_i)$ dado $\boldsymbol{\beta}$ baseada nos erros $\epsilon_1,\ldots,\epsilon_n$. A verossimilhança marginal PR de $\boldsymbol{\beta}$ é escrita como:

$$L_{PR}(\boldsymbol{\beta}) = \prod_{i=1}^{n} f_{i-1,\boldsymbol{\beta}}\left(y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta})\right).$$

Como na Seção 2.2.2, estima-se $\boldsymbol{\beta}$ pela maximização da verossimilhança PR marginal, ou, equivalentemente, a log-verossimilhança PR marginal $\ell_{PR}(\boldsymbol{\beta}) = \log\left(L_{PR}(\boldsymbol{\beta})\right)$.

Para sua implementação, são necessárias as definições de algumas medidas:

- Para os cálculos no algoritmo PR, é preciso definir o suporte da distribuição da variável de mistura U. Após um pequeno estudo de simulação com o objetivo de identificar um intervalo de $\mathscr{U}$ tal que obtivéssemos boas estimativas de $\boldsymbol{\beta}$, sugerimos $\mathscr{U} = [U_{\min}; U_{\max}]$, em que $U_{\min} = 5 \times 10^{-4}$ e $U_{\max} = 100$.
- Para inicializar o algoritmo, deve se definir uma f.d.p. para $\psi_0(u)$. Desta forma, assim como em Martin e Han (2016), utilizamos $\psi_0(u)$ como sendo a f.d.p. de uma distribuição uniforme no intervalo $[U_{\min}; U_{\max}]$.
- Para a dependência do algoritmo PR em relação a ordem dos dados, utilizamos 25 permutações, como sugerido em Martin e Han (2016).

Algoritmo Híbrido PR-EM

A extensão do algoritmo híbrido PR-EM para os modelos de regressão não lineares em que o parâmetro de escala σ^2 é conhecido e igual a 1 é proposto neste trabalho, generalizando então, o algoritmo PR-EM discutido em (MARTIN; HAN, 2016).

A ideia da sua construção é associar novamente o algoritmo PR (utilizando na estimação de $\psi(u)$) ao algoritmo EM (aplicado para a estimação de máxima verossimilhança de $\boldsymbol{\beta}$). Assim, reescrevendo $L_{PR}(\boldsymbol{\beta})$, tem-se

$$L_{PR}(\boldsymbol{\beta}) = \prod_{i=1}^{n} f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right) \times \frac{\phi\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right);0,u^{-1}\right)}{\phi\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right);0,u^{-1}\right)}.$$

Calculando a log-verrossimilhança, tem-se:

$$\ell_{PR}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\left[\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)\right] - \sum_{i=1}^{n} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta},(y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}\right]. \tag{2.8}$$

Assim, redefinindo a densidade em (2.5) para o cálculo da esperança na equação (2.8), tem-se:

$$\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u) = \frac{\psi_{i-1,\boldsymbol{\beta}^{(t)}}(u)\phi\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)},(y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}, \tag{2.9}$$

em que $\boldsymbol{\beta}^{(t)}$ é uma atualização da estimativa obtida pelo algoritmo no passo t. Assim, para o Passo E do algoritmo EM, deve-se obter a função $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$, definida por:

$$Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) + Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right),$$

em que

$$Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = \sum_{i=1}^{n} \int_{\mathscr{U}} \log\left[\phi\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right);0,u^{-1}\right)\right] \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du$$

e

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\sum_{i=1}^{n} \int_{\mathscr{U}} \log\left[\frac{\phi\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta},(y_i-g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}\right] \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du.$$

Reescrevendo $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, tem-se:

$$Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\frac{1}{2}\sum_{i=1}^{n}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)^2 \hat{\omega}_i + c\left(\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}^{(t)}\right),$$

em que

$$\hat{\omega}_i = \int_{\mathscr{U}} u\ \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du, \tag{2.10}$$

Neste caso, $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ depende apenas de $\psi_{i-1,\boldsymbol{\beta}^{(t)}}$, de $y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)$, da estimativa $\boldsymbol{\beta}^{(t)}$ e de $c\left(\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}^{(t)}\right)$, uma função que independe de $\boldsymbol{\beta}$.

Ao se considerar a função $Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ como a soma de duas funções $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ e $Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, produzimos um processo de maximizações somente na função $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$. Essa abordagem se torna interessante, pois

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) \xrightarrow{\boldsymbol{\beta}^{(t)}\to\boldsymbol{\beta}} 0.$$

Além do mais, escolher a densidade dada em (2.9), para a obtenção da função $Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ reduz o processo de maximização de $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ em uma sequência de soluções de mínimos quadrados de um modelo não linear ponderado pela matriz de pesos $\widehat{\boldsymbol{\Omega}}$, formada pelos pesos $\widehat{\omega}_1,\ldots,\widehat{\omega}_n$ dado em (2.10).

Considere

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\sum_{i=1}^{n} \int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta},(y_i-g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}\right] \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du.$$

Reescrevendo $Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$, tem-se:

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) = -\sum_{i=1}^{n} \int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta},(y_i-g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}\right] \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du + nD_n\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right),$$

em que $D_n\left(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}\right)$ é expressa por

$$D_n\left(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}\right) = \frac{1}{n}\sum_{i=1}^{n} \log\left[\frac{f_{i-1,\boldsymbol{\beta}}(y_i - g(x_i, \boldsymbol{\beta})}{f_{i-1,\boldsymbol{\beta}^{(t)}}(y_i - g(x_i, \boldsymbol{\beta})}\right].$$

Desta forma,

$$Q_2\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) = -\sum_{i=1}^{n}\int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}(y_i - g(\boldsymbol{x}_i;\boldsymbol{\beta}))}}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du + nD_n\left(\boldsymbol{\beta}^{(t)}, \boldsymbol{\beta}^{(t)}\right),$$

em que $D_n\left(\boldsymbol{\beta}^{(t)}, \boldsymbol{\beta}^{(t)}\right)$ é expressa por

$$D_n\left(\boldsymbol{\beta}^{(t)}, \boldsymbol{\beta}^{(t)}\right) = \frac{1}{n}\sum_{i=1}^{n} \log\left[\frac{f_{i-1,\boldsymbol{\beta}^{(t)}}(y_i - g(x_i, \boldsymbol{\beta})}{f_{i-1,\boldsymbol{\beta}^{(t)}}(y_i - g(x_i, \boldsymbol{\beta})}\right] = 0.$$

Calculando $Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_2\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right)$, tem-se:

$$\begin{aligned}
Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_2\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) &= -\sum_{i=1}^{n}\int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du \\
&+ \sum_{i=1}^{n}\int_{\mathscr{U}} \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du \\
&+ nD_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)} \\
&= \sum_{i=1}^{n}\int_{\mathscr{U}} \left\{\log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}^{(t)}\right)\right)}\right]\right. \\
&\left. - \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right]\right\}\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du \\
&+ nD_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)} \\
&= \sum_{i=1}^{n}\int_{\mathscr{U}} I\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du + nD_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}),
\end{aligned}$$

em que

$$I = \left\{\log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}^{(t)}\right)\right)}\right] - \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right]\right\}$$

Pode-se reescrever I na seguinte forma:

$$
\begin{aligned}
I &= \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}^{(t)}\right)\right)}\frac{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)}\frac{\psi_{i,\boldsymbol{\beta}^{(t)}}(u)}{\psi_{i,\boldsymbol{\beta}^{(t)}}(u)}\right] \\
&= \log\left[\frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)\psi_{i,\boldsymbol{\beta}^{(t)}}(u)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}^{(t)}\right)\right)}\frac{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)\psi_{i,\boldsymbol{\beta}^{(t)}}(u)}\right] \\
&= \log\left[\frac{\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)}{k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u)}\right],
\end{aligned}
$$

em que

$$
k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u) = \frac{\phi\left(\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right);0,u^{-1}\right)\psi_{i,\boldsymbol{\beta}^{(t)}}(u)}{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}.
$$

Desta forma,

$$
Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_2\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) = \sum_{i=1}^{n}\int_{\mathscr{U}}\log\left[\frac{\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)}{k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u)}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du + nD_n\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right).
$$

Pela desigualdade Kullbeck-Leibler (EGUCHI; COPAS, 2006), tem-se que:

$$
\log\left[\frac{\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)}{k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u)}\right]\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)du \geq 0.
$$

Sendo assim,

$$
\begin{aligned}
Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) &= Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_1\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) + Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_2\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) \\
&\geq Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_1\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) + nD_n\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right)
\end{aligned}
$$

Reescrevendo $D_n\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right)$, obtém-se:

$$
\begin{aligned}
D_n\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right) &= \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\frac{f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right] - \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}{f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)}\right],
\end{aligned}
$$

em que $f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)$ denota a verdadeira distribuição de $\boldsymbol{Y}$.

Sendo assim, $D_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) \geq 0$, pois:

- $D_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) = 0$ se $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}$;
- $D_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) > 0$ se $\boldsymbol{\beta}$ assumir o verdadeiro valor, pois pode-se escolher $\boldsymbol{\beta}$ tal que $Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) > Q\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right)$.

Note que, pode-se escolher $\boldsymbol{\beta}$, tal que, $f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right) \approx f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)$, o que torna o segundo termo de $D_n(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}$ igual a 0. Assim, quando $\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}$, $f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)$ se aproxima de $f_{i-1,\boldsymbol{\beta}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right)$.

Desta forma, na convergência de $\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}$,

$$f_{i-1,\boldsymbol{\beta}^{(t)}}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right) \approx f^*\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}\right)\right).$$

Como consequência,

$$D_n\left(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}\right) \xrightarrow{\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}} 0.$$

Desta forma,

$$Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right) \geq Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) - Q_1\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right).$$

Escolhendo $\boldsymbol{\beta}$, tal que $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) \geq Q_1\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right)$, tem-se:

$$Q\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) \geq Q\left(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}\right).$$

Provando assim, a propriedade de ascendência do algoritmo.

Pode-se observar também que, na convergência de $\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}$, $\log\left[\frac{\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u)}{k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u)}\right] \to 0$. Isso ocorre pois,

$$\psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u) \xrightarrow{\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}} k_{i,\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}}(u).$$

Desta forma,

$$Q_2\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right) \xrightarrow{\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}} 0.$$

Algoritmo PR-EM

A seguir, serão apresentados os passos para a estimação do vetor de parâmetros $\boldsymbol{\beta}$.

Dado $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$ e $(\mathscr{U}, \psi_0, w_1, \ldots, w_n)$, referente ao algoritmo PR, tem-se:

Passo E: Calcule os pesos $\hat{\omega}_1, \ldots, \hat{\omega}_n$ utilizando o algoritmo PR com os erros $y_1 - g\left(\boldsymbol{x}_1\boldsymbol{\beta}^{(t)}\right)$, $\ldots$, $y_n - g\left(\boldsymbol{x}_n\boldsymbol{\beta}^{(t)}\right)$.

Passo M: Atualiza-se $\boldsymbol{\beta}^{(t)}$ pela maximização $Q_1\left(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}\right)$ em $\boldsymbol{\beta}$, o que leva a minimização de

$$\sum_{i=1}^{n}\left(y_i - g\left(\boldsymbol{x}_i;\boldsymbol{\beta}^{(t)}\right)\right)^2\hat{\omega}_i,$$

em que $\hat{\omega}_1, \ldots, \hat{\omega}_n$ são os pesos em (2.6).

Os passos E e M serão repetidos até que a distância entre duas avaliações sucessivas de $\boldsymbol{\beta}$, dado por:

$$||\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}|| < \delta,$$

tal que δ seja suficientemente pequeno.

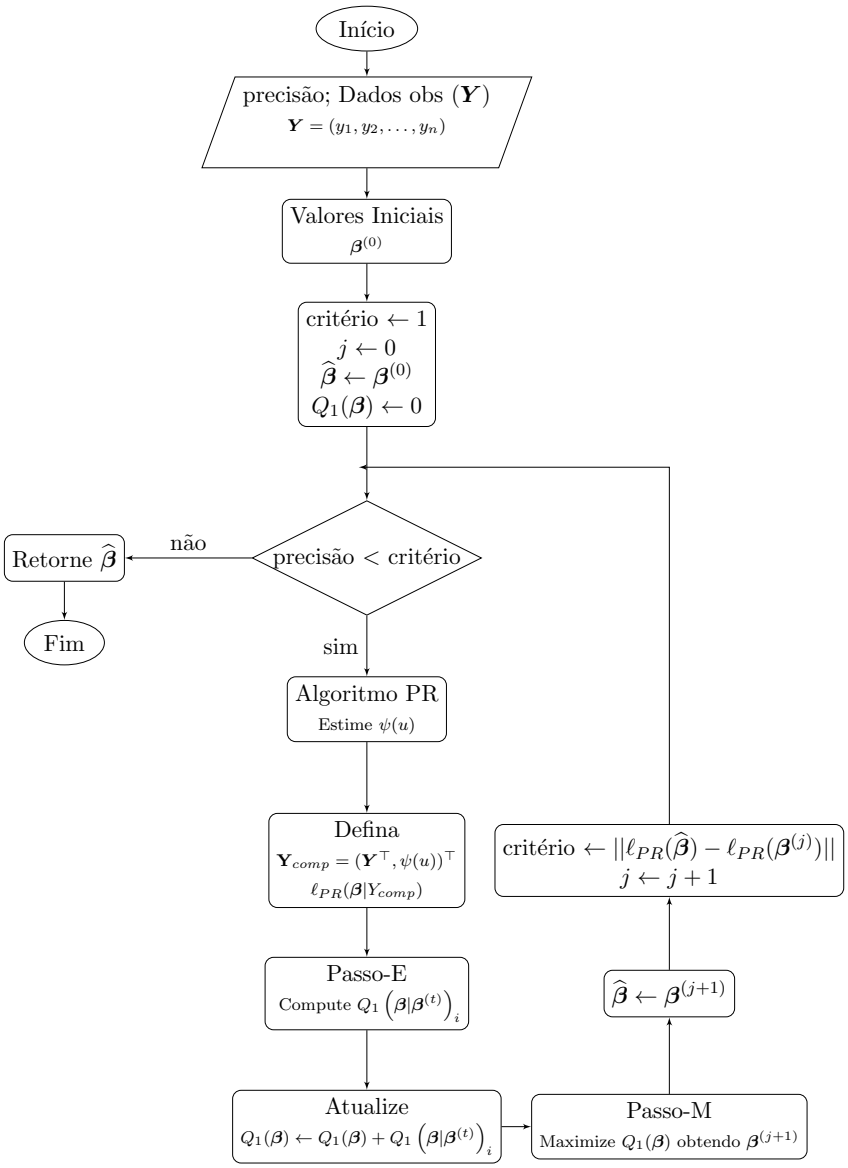
Sugere-se como valores iniciais para $\boldsymbol{\beta}^{(0)}$ as estimativas de máxima verossimilhança $\boldsymbol{\beta}^{\top}$ do modelo de regressão não linear normal. Tendo em vista facilitar o entendimento do algoritmo proposto nessa dissertação, a Figura 2.1 apresenta o fluxograma com todos os passos necessários para a implementação do algoritmo PR-EM

2.4 Estimação dos erros padrão dos estimadores do modelo

O cálculo do erro padrão é de grande importância para avaliar a variabilidade dos estimadores dos parâmetros do modelo proposto. Nesta seção, serão utilizadas duas metodologias para a estimação dos erros padrão dos estimadores. Primeiramente é apresentado o método via *boostrap* apresentado por McLachlan e Krishnan (2007) é interessante buscar uma alternativa para esse método. Em seguida, apresenta-se a metodologia desenvolvida por Louis (1982) que consiste em aproximar a matriz de variâncias e covariâncias por meio da matriz de informação empírica.

A primeira é o método via *boostrap*, porém, dado seu alto custo computacional,

Figura 2.1: Fluxograma do algoritmo PR-EM



2.4.1 Erros padrão via *bootstrap*

Proposto inicialmente por Efron (1979) e é, geralmente, utilizado para estimar a distribuição das estatísticas de interesse, que em algumas situações não podem ser obtidas analiticamente ou assintoticamente.

A ideia principal do *bootstrap* é considerar os dados amostrais como sendo os dados populacionais e gerar amostras (com reposição) desses dados amostrais como sendo amostras da população. Repetir esse procedimento até obter B subamostras e, para cada uma delas, calcular a quantidade de interesse. Finalmente, utilizar os B valores da quantidade de interesse que foram calculadas para estimar sua distribuição empírica.

Segundo Moore et al. (2006), os métodos de reamostragem (*bootstrap*, *Jackknife*, entre outros), permitem quantificar a incerteza calculando os erros padrão, intervalos de confiança, assim como a realização de testes de hipóteses. Sua utilização, em algumas situações, pode não exigir a suposição de uma distribuição para os dados e geralmente fornece respostas mais precisas do que os métodos tradicionais.

Seja uma amostra aleatória $\boldsymbol{y} = (y_1, \ldots, y_n)$ da variável aleatória Y, com f.d.p. e f.d.a., denotadas por $f(y)$ e $F(y)$, respectivamente. O método *bootstrap* pode ser paramétrico ou não-paramétrico. O *bootstrap* não paramétrico considera que a f.d.a. $F(\cdot)$ é desconhecida e que pode ser estimada pela função de distribuição empírica,

$$\widehat{F}(y) = \frac{\sum_{i=1}^{n} \mathbb{I}_{\{y_i \leq y\}}}{n},\ i = 1, \ldots, n,$$

em que

$$\mathbb{I}_{\{y_i \leq y\}} = \begin{cases} 1, & \text{se } y_i \leq y, \\ 0, & \text{caso contrário} \end{cases}$$

e que atribui probabilidade $\frac{1}{n}$ em cada valor amostral y_i, $i = 1, \ldots, n$. No caso não-paramétrico é supõe-se apenas que $(y_1, \ldots, y_n)$ são geradas de forma independente. No *bootstrap* paramétrico, supõe-se que a f.d.a. da v.a. Y é dada por $F(\boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor de parâmetros (podendo ser um escalar).

Considere o modelo de regressão,

$$y_i = g(\boldsymbol{x}_i, \boldsymbol{\beta}) + \epsilon_i,\ i = 1, \ldots, n, \tag{2.11}$$

em que $\epsilon_i \sim MEN(0, \sigma^2 = 1, \Psi)$ e $g(\boldsymbol{x}_i, \boldsymbol{\beta})$ é uma função diferenciável.

Seja $\boldsymbol{\beta}$ o vetor de parâmetros do modelo. Para estimar os erros padrão do estimador $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^\top$, o uso do *bootstrap* não paramétrico é o mais indicado, pois, como apresentado na Seção 2.3, a distribuição da mistura U é considerada desconhecida, desta forma, não é possível obter analiticamente, a função de probabilidade dos erros $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. Assim, se torna inviável a utilização do método *bootstrap* paramétrico.

Os passos do algoritmo *boostrap* para estimação dos erros padrão dos estimadores do modelo são descritos a seguir:

Passo 1: Reamostre com reposição, M pares independentes $(\boldsymbol{y}_1^*, \boldsymbol{x}_1^*), \ldots, (\boldsymbol{y}_B^*, \boldsymbol{x}_B^*)$, cada amostra constituindo de n valores retirados do valores amostrais $(\boldsymbol{y}, \boldsymbol{x})$;

Passo 2: Ajuste o modelo de regressão dado em (2.11) por meio do algoritmo PR-EM para cada par $(\boldsymbol{y}_b^*, \boldsymbol{x}_b^*)$ e estime $\widehat{\boldsymbol{\beta}}$, que será denotado por $\widehat{\boldsymbol{\beta}}^{(b)}$, $b = 1, \ldots, B$;

Passo 3: Estime o erro padrão de $\widehat{\beta}_i$, $i = 1, \ldots, p$, por

$$\widehat{\mathrm{ep}}_B\left(\hat{\beta}_i\right) = \left\{ \sum_{b=1}^{B} \frac{\left[\hat{\beta}_i^{(b)} - \hat{\beta}_i^{*}\right]^2}{B-1} \right\}^{1/2},$$

em que $\hat{\beta}_i^{*} = \sum_{b=1}^{B} \frac{\hat{\beta}_i^{(b)}}{B}$.

Efron e Tibshirani (1994) sugerem valores de B entre 50 e 200.

2.4.2 Matriz de informação empírica

Como alternativa ao desenvolvimento analítico dos erros padrão dos estimadores do modelo com o uso do algoritmo PR-EM apresentado na Seção 2.2.2, propõem-se o uso do método *Bootstrap*. Com um alto custo computacional, métodos alternativos foram propostos na literatura. Meilijson (1989) propõe um método para o cálculo da matriz de variâncias e covariâncias para estimadores de máxima de verossimilhança quando a matriz de informação observada não possui uma forma analítica tratável ou é inviável seu cálculo, para mais detalhes veja Meilijson (1989) e Lin (2009). O método consiste em aproximar a matriz de informação através da matriz de informação empírica, dada por:

$$I_e = \sum_{i=1}^{n} \boldsymbol{v}\left(y_i|\boldsymbol{\beta}\right) \boldsymbol{v}^{\top}\left(y_i|\boldsymbol{\beta}\right) - \frac{1}{n} \boldsymbol{V}\left(\boldsymbol{y}|\boldsymbol{\beta}\right) \boldsymbol{V}^{\top}\left(\boldsymbol{y}|\boldsymbol{\beta}\right), \tag{2.12}$$

em que $\boldsymbol{V}\left(\boldsymbol{y}|\boldsymbol{\beta}\right) = \sum_{i=1}^{n} \boldsymbol{v}\left(y_i|\boldsymbol{\beta}\right)$. Considerando o resultado obtido por Louis (1982), o vetor escore para a observação y_i, $i = 1, \ldots, n$, pode ser obtido da seguinte forma:

$$\boldsymbol{v}\left(y_i|\boldsymbol{\beta}\right) = \frac{\partial \ell\left(\boldsymbol{\beta}|y_i\right)}{\partial \boldsymbol{\beta}} = E\left[\left. \frac{\partial \ell_c\left(\boldsymbol{\beta}|y_i, U\right)}{\partial \boldsymbol{\beta}} \right| y_i, \boldsymbol{\beta} \right],$$

em que $\ell_c(\cdot)$ é a log-verossimilhança completa.

Sob condições de regularidade (COX; HINKLEY, 1979), tem-se

$$\boldsymbol{v}\left(y_i|\boldsymbol{\beta}\right) = \frac{\partial Q\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)}{\partial \boldsymbol{\beta}}.$$

Assim, substituindo as estimativas de $\boldsymbol{\beta}$, denotadas por $\hat{\boldsymbol{\beta}}$, em (2.12), a matriz de informação empírica é reduzida a

$$I_e = \sum_{i=1}^{n} \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i^{\top},$$

em que $\hat{\boldsymbol{v}}_i$ representa vetor escore para a observação y_i, $i = 1, \ldots, n$, dado por

$$\hat{\boldsymbol{v}}_i = \left.\frac{\partial Q_{i,1}\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \left.\frac{\partial Q_{i,2}\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}. \tag{2.13}$$

em que $Q_{i,1}\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)$ e $Q_{i,2}\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)$ são as funções $Q_1\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)$ e $Q_2\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}\right)$ avaliadas na observação y_i, $i = 1, \ldots, n$.

Sendo assim,

$$\frac{\partial Q_{i,1}\left(\beta_j|\hat{\boldsymbol{\beta}}\right)}{\partial \beta_j} = \widehat{\omega}_i \left(y_i - g\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right)\right) \frac{\partial g\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right)}{\partial \beta_j}$$

e

$$\frac{\partial Q_{i,2}\left(\beta_j|\hat{\boldsymbol{\beta}}\right)}{\partial \beta_j} = -\int_{\mathscr{U}} \left(y_i - g\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right)\right) u \frac{\partial g\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right)}{\partial \beta_j} \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u) du - \int_{\mathscr{U}} \frac{1}{f_{i-1,\widehat{\boldsymbol{\beta}}}} \frac{\partial f_{i-1,\widehat{\boldsymbol{\beta}}}}{\partial \beta_j} \psi^{\mathrm{B}}_{i,\boldsymbol{\beta}^{(t)}}(u) du,$$

em que

$$\frac{\partial f_{i-1,\widehat{\boldsymbol{\beta}}}}{\partial \beta_j} = \int_{\mathscr{U}} \frac{\partial \phi\left(\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right)}{\partial \beta_j} \psi_{i-1}(u) + \frac{\partial \psi_{i-1}(u)}{\partial \beta_j} \phi\left(\left(y_i - \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right) du$$

e

$$\begin{aligned}\frac{\partial \psi_{i-1}(u)}{\partial \beta_j} = (1 - w_{i-1}) \frac{\partial \psi_{i-2}(u)}{\partial \beta_j} + \psi_{i-1}(u) \Bigg[& \frac{\partial \phi\left(\left(y_{i-1} - \boldsymbol{x}_{i-1}^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right)}{\partial \beta_j} \frac{\psi_{i-2}(u)}{f_{i-2,\widehat{\boldsymbol{\beta}}}} \\ & + \frac{\partial \psi_{i-2}(u)}{\partial \beta_j} \frac{\phi\left(\left(y_{i-1} - \boldsymbol{x}_{i-1}^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right)}{f_{i-2,\widehat{\boldsymbol{\beta}}}} - \frac{\partial f_{i-2,\widehat{\boldsymbol{\beta}}}}{\partial \beta_j} \frac{\phi\left(\left(y_{i-1} - \boldsymbol{x}_{i-1}^{\intercal}\boldsymbol{\beta}\right); 0, u^{-1}\right) \psi_{i-2}(u)}{f^2_{i-2,\widehat{\boldsymbol{\beta}}}} \Bigg].\end{aligned}$$

Este procedimento também foi utilizado por Wang e Lin (2015) em modelagem de

clusters sob distribuição *skew-t*, Garay et al. (2015) em modelos de regressão linear sob mistura de escala normal para dados censurados e Garay et al. (2016) no contexto dos modelos de regressão não lineares sob mistura de escala normal para dados censurados.

CAPÍTULO 3

Simulações e Aplicação

3.1 Estudos de Simulação

Nesta seção serão apresentados estudos de simulação de Monte Carlo com os seguintes objetivos:

i) avaliar o comportamento dos estimadores obtidos pelo o algoritmo PR-EM;

ii) avaliação a robustez dos estimadores obtidos por meio do algoritmo PR-EM quando há observações atípicas.

Os indicadores utilizados nesta seção para a avaliação serão o erro quadrático médio empírico (EQM-E), viés relativo empírico (VR-E) e a mudança relativa empírica (MR-E), definidos por:

$$\text{EQM-E}\left(\hat{\beta}_j\right) = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\beta}_j^{(i)} - \beta_j\right)^2,$$

$$\text{VR-E}\left(\hat{\beta}_j\right) = \frac{1}{M}\sum_{i=1}^{M}\frac{|\hat{\beta}_j^{(i)} - \beta_j|}{\beta_j}$$

e

$$\text{MR-E}\left(\hat{\beta}_j(\nabla)\right) = \frac{|\hat{\beta}_j(\nabla) - \hat{\beta}_j|}{\hat{\beta}_j},$$

em que

- M é o número de repetições de Monte Carlo;
- $\hat{\beta}_j^{(i)}$ é a estimativa do parâmetro β_j para a i-ésima réplica de Monte Carlo;
- $\hat{\beta}_j(\nabla)$ a estimativa do parâmetro β_j sob observações atípicas para algum ∇;
- β_j é o verdadeiro valor do parâmetro;
- A estimativa de β_j é dada por:

$$\hat{\beta}_j = \frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_j^{(i)}.$$

Durante todos os estudos de simulação foi considerado as seguintes modelos não lineares:

$$y_i = \beta_1 + e^{\beta_2 x_i} + \epsilon_i,\ i = 1, 2, \ldots, n, \tag{3.1}$$

e

$$y_i = e^{\beta_1 + \beta_2 x_i} + \epsilon_i,\ i = 1, 2, \ldots, n. \tag{3.2}$$

Em todos os estudos de simulação os valores da variável explicativa x_i foram geradas a partir de uma distribuição $\mathcal{U}(0, 1)$ e mantidos fixos. Foram realizadas 5000 réplicas de Monte Carlo e os códigos foram implementados na linguagem de programação R.

3.1.1 Estudo de Simulação 1

Nesse primeiro estudo foram avaliados os desempenhos dos estimadores dos parâmetros obtidos com o algoritmo PR-EM para MRNL-MEN com os preditores apresentados em (3.1) e (3.2) sob quatro casos. Para cada um dos preditores, assume-se uma distribuição para os erros $\epsilon_i, i = 1, \ldots, n$ que são:

1. $\epsilon_i \sim N(0, 1)$;
2. $\epsilon_i \sim t(0, \nu = 5)$;
3. $\epsilon_i \sim t(0, \nu = 3)$;
4. $\epsilon_i \sim Slash(0, \nu = 2)$.

Em todos os casos, foi fixado os verdadeiros valores dos parâmetros como sendo $\beta_1 = 3$ e $\beta_2 = 1$. Sendo os tamanhos amostrais $n = 15, 30, 50, 75, 100, 200$ e 500.

Na Tabela 3.1 são apresentados os resultados numéricos referentes as estimativas dos parâmetros para o modelo de regressão apresentado em (3.1) sob erros normais, t de Student com 3 e 5 graus de liberade e Slash com 2 graus de liberdade obtidas com o algoritmos PR-EM. Avaliando as estimativas obtidas com o algoritmo PR-EM pode-se observar que os EQM-E dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ decaem com o aumento do tamanho amostral e o mesmo comportamento é observado para as estimativas de VR-E dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$. Na Tabela 3.2 são apresentados os resultados numéricos referentes as estimativas dos parâmetros para o modelo de regressão apresentado em (3.2) sob erros normais, t de Student com 3 e 5 graus de liberade e Slash com 2 graus de liberdade obtidas com o algoritmos PR-EM. Avaliando as estimativas obtidas com o algoritmo PR-EM pode-se observar que os EQM-E e VR-E dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ decaem com o aumento de n.

Com os resultados das simulações é possível observar que as estimativas de EQM e viés relativo dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ obtidos através do algoritmo PR-EM, em geral, decaem com o aumento do tamanho amostral, indicando uma convergência do EQM para zero. O mesmo comportamento, de forma geral, é observado nas estimativas de viés relativo. Desta forma, os resultados indicam, empiricamente, que os estimadores obtidos por meio algoritmo PR-EM apresentam estimadores assintoticamente não viesados e consistentes.

Tabela 3.1: Erro quadrático médio e viés relativo dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ obtidos com o algoritmo PR-EM para diferentes tamanhos amostrais para o modelo (3.1) para os diferentes casos.

n	Normal				t(3)			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E
15	0,50675	0,01639	1,76969	0,31659	0,41991	0,00954	1,49206	0,18418
30	0,15695	0,00733	0,29639	0,06371	0,37091	0,00664	1,39950	0,09981
50	0,09077	0,00198	0,06159	0,02335	0,10268	0,00140	0,10538	0,03151
75	0,06406	0,00157	0,04788	0,01737	0,06608	0,00125	0,04975	0,02315
100	0,04479	0,00110	0,02808	0,01550	0,06241	0,00115	0,04210	0,01613
200	0,02083	0,00041	0,01269	0,00394	0,05376	0,00103	0,02974	0,00963
500	0,00775	0,00034	0,00555	0,00369	0,00907	0,00076	0,01139	0,00705
n	t(3)				Slash(2)			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E
15	0,35335	0,01651	4,64761	0,34342	0,35335	0,01651	4,64761	0,34342
30	0,21647	0,01065	1,49443	0,14628	0,21647	0,01065	1,49443	0,14628
50	0,16348	0,00559	0,48730	0,06714	0,16348	0,00559	0,48730	0,06714
75	0,07341	0,00139	0,05912	0,02778	0,07341	0,00139	0,05912	0,02778
100	0,06406	0,00125	0,04380	0,02072	0,06406	0,00125	0,04380	0,02072
200	0,03226	0,00105	0,01821	0,00503	0,03226	0,00105	0,01821	0,00503
500	0,01224	0,00039	0,00712	0,00364	0,01224	0,00039	0,00712	0,00364

3.1.2 Estudo de Simulação 2

Neste estudo foram avaliados o desempenho dos estimadores dos parâmetros obtidos com o algoritmo PR-EM para MRNL com preditor apresentado em (3.1) na presença de observações atípicas. Fez-se $\epsilon_i \sim N(0, \sigma^2 = 1), i = 1, \ldots, n$ e, em cada réplica de Monte Carlo, uma perturbação no maior valor para torná-la atípica foi dada foi inserido da seguinte forma

$$y_{\max} = y_{\max} + \nabla * V(\boldsymbol{y}),$$

em que $y_{\max}$ representa o maior valor observado na amostra $\boldsymbol{y}$, $V(\boldsymbol{y})$ a variância amostral de $\boldsymbol{y}$ e $\nabla = 1, 2 \ldots, 10$.

As estimativas dos parâmetros foram obtidas de duas formas:

1. através do modelo não linear clássico sobre erros normais, que será denotado por MRNL-N;

2. por meio do algoritmo PR-EM nos MRNL-MEN.

Tabela 3.2: Erro quadrático médio e viés relativo dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ obtidos com o algoritmo PR-EM para diferentes tamanhos amostrais para o modelo (3.2) para os diferentes casos.

n	Normal				t(3)			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E
15	0,00085	0,00042	0,00215	0,07177	0,00079	0,00062	0,00136	0,00194
30	0,00021	0,00040	0,00047	0,06371	0,00035	0,00060	0,00071	0,00147
50	0,00016	0,00038	0,00035	0,02335	0,00027	0,00054	0,00055	0,00137
75	0,00011	0,00037	0,00016	0,01737	0,00021	0,00051	0,00037	0,00129
100	0,00006	0,00035	0,00013	0,00117	0,00011	0,00047	0,00024	0,00124
200	0,00003	0,00033	0,00007	0,00096	0,00005	0,00046	0,00012	0,00119
500	0,00001	0,00032	0,00003	0,00095	0,00002	0,00040	0,00005	0,00107
n	t(3)				Slash(2)			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E	EQM-E	VR-E
15	0,00079	0,00062	0,00136	0,00147	0,00099	0,00059	0,00168	0,00179
30	0,00035	0,00060	0,00071	0,00136	0,00042	0,00056	0,00086	0,00167
50	0,00026	0,00034	0,00059	0,00085	0,00023	0,00053	0,00046	0,00159
75	0,00022	0,00030	0,00049	0,00071	0,00018	0,00047	0,00038	0,00137
100	0,00014	0,00025	0,00032	0,00062	0,00012	0,00045	0,00027	0,00114
200	0,00007	0,00023	0,01821	0,00051	0,00007	0,00061	0,00014	0,00103
500	0,00003	0,00018	0,00712	0,00047	0,00003	0,00029	0,00006	0,00044

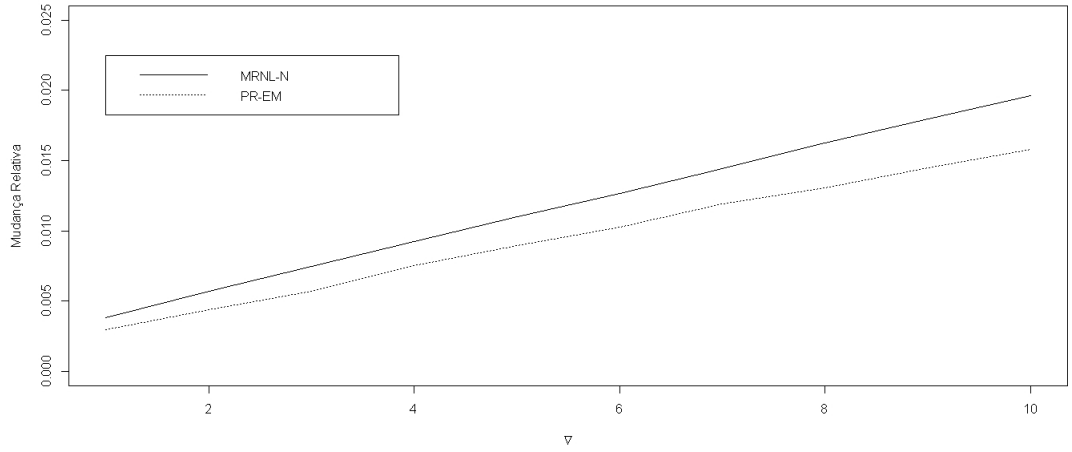
Em todos os cenários, fixou-se os valores verdadeiros dos parâmetros como sendo $\beta_1 = 3$, $\beta_2 = 1$ e o tamanho amostral $n = 100$.

A Tabela 3.3 são apresentados os resultados numéricos referentes as estimativas empíricas de MR-E dos parâmetros do modelo de regressão não linear. Analisando os estimadores do parâmetro β_1, tem-se que, para o MRNL-N, as estimativas de MR-E aumentam com o aumento do valor de ∇. As estimativas de MR-E obtidas por meio do algoritmo PR-EM também apresentam o comportamento crescente conforme o aumento de ∇.

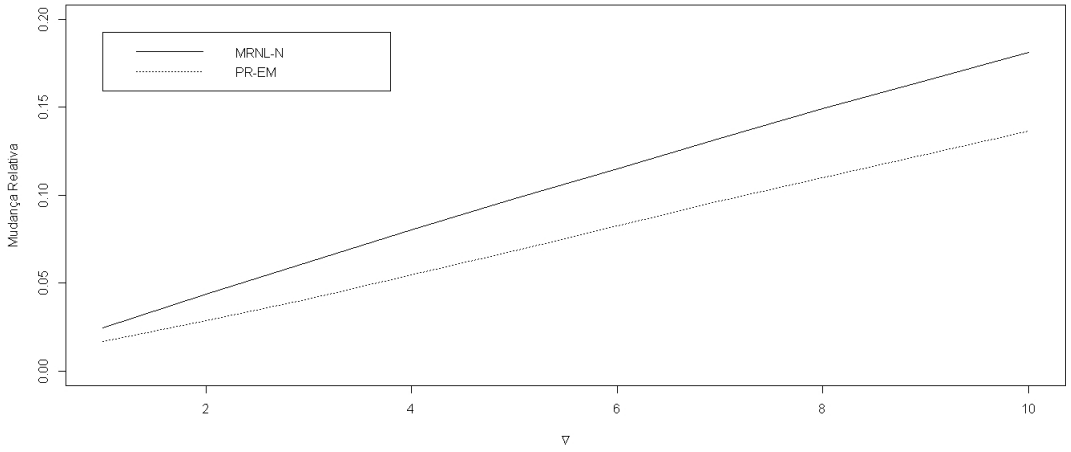
No caso dos estimadores do parâmetro β_2, observa-se que, para as estimativas de MR-E do MRNL-N há um crescimento acentuado desses valores com o aumento de ∇. As estimativas de MR-E obtidas através do algoritmo PR-EM também crescem com aumento de ∇.

A Figura 3.1 apresenta os gráficos das mudanças relativas das estimativas para os diferentes métodos de estimação dos parâmetros e valores de ∇. Pode-se observar que as estimativas de mudança relativa, para ambos os estimadores, obtidas através do algoritmo PR-EM, são inferiores que as estimativas obtidas por meio do MRNL-N. Desta forma, os estimadores obtidos por meio do algoritmo PR-EM apresentam uma menor influência em suas estimativas na presença de valores atípicos que os estimadores obtidos através do MRNL-N.

Figura 3.1: Mudança Relativa dos estimadores dos parâmetros dos modelos em estudo.



(a) β_1



(b) β_2

Tabela 3.3: Mudança relativa dos estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ obtidos por meio do MRNL-N e com o algoritmo PR-EM para o modelo (3.1).

MR-E $\left(\hat{\beta}_j(\nabla)\right)$	Modelos	∇									
		1	2	3	4	5	6	7	8	9	10
$MR(\beta_1)$	MRNL-N	0.003801	0.00569	0.00745	0.00924	0.01103	0.01269	0.01444	0.01625	0.01797	0.01962
	PR-EM	0.00294	0.00437	0.00571	0.00754	0.00865	0,01026	0,01191	0,01309	0,01450	0,01579
$MR(\beta_2)$	MRNL-N	0.02447	0.04369	0.06210	0.08026	0.09811	0.11511	0.13230	0.14912	0.16543	0.18122
	PR-EM	0,01679	0,02875	0,04087	0,05494	0,06855	0,08247	0,09683	0,11008	0,12321	0,13644

3.2 Aplicação

Nesta seção será apresentado um conjunto de dados reais em que a metodologia estudada é aplicada. Esse conjunto de dados é referente ao índice de prognóstico para recuperação a longo prazo.

Índice de prognóstico para recuperação a longo prazo

Considere o conjunto de dados descrito em Kutner et al. (2004), em que o autor tem o interesse de encontrar uma relação entre o logaritmo do índice de prognóstico para recuperação a longo prazo, y e o tempo em que o paciente foi hospitalizado, x (dias), em amostra de 15 pacientes após serem expostos a descargas elétricas. Na Tabela 3.4 encontram-se os resultados de algumas medidas descritivas. Nota-se que a variável resposta, y, assume valores entre 1,39 e 3,99. Já a variável explicativa, x, assume valores entre 2 e 65. Observa-se que o logaritmo do índice de prognóstico é, em média, de 2,87 e o tempo de internação, em média, é de 30,73 dias. Nota-se também que os valores da variável resposta, y, possuem menor variabilidade que os valores da variável explicativa, x, uma vez que o coeficiente de variação de y, 28,41 %, é bem menor que o coeficiente de variação de x, 68,29 %.

Tabela 3.4: Medidas resumo das variáveis y e x.

Medida	y	x
Mínimo	1,39	2
Média	2,87	30,73
Mediana	2,89	31
Máximo	3,99	65
Desvio padrão	0,6651	20,9880
Coeficiente de Variação (%)	28,41	68,29

As Figuras 3.2 e 3.3, apresentam os boxplots para as variáveis y e x, respectivamente. Pode-se perceber que, uma simetria nas observações da variável x e uma leve assimetria à esquerda na variável y. A Figura 3.4 apresenta o gráfico de dispersão das variáveis y e x. Observa-se que com o aumento do tempo em que o paciente foi hospitalizado menor é o logaritmo índice de prognóstico.

Com base no modelo em Kutner et al. (2004), propõe-se o seguinte modelo para os dados:

$$y_i = \alpha + \beta x_i + \epsilon_i,\ i = 1, \ldots, 15, \tag{3.3}$$

em que $\epsilon_i \sim MEN(0, 1, \Psi)$.

A Tabela 3.5 apresenta as estimativas dos parâmetros obtidas com o algoritmo PR-EM e com o ajuste de um modelo linear com erros sob distribuição normal para o modelo em

Figura 3.2: Boxplot do logaritmo índice de prognóstico para recuperação a longo prazo.

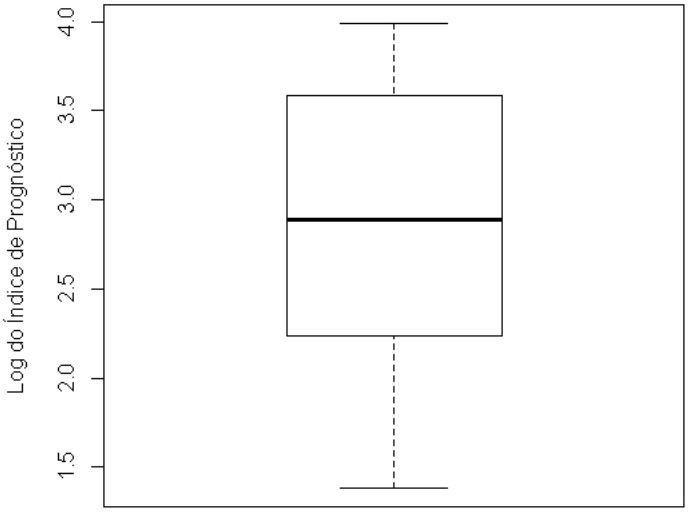
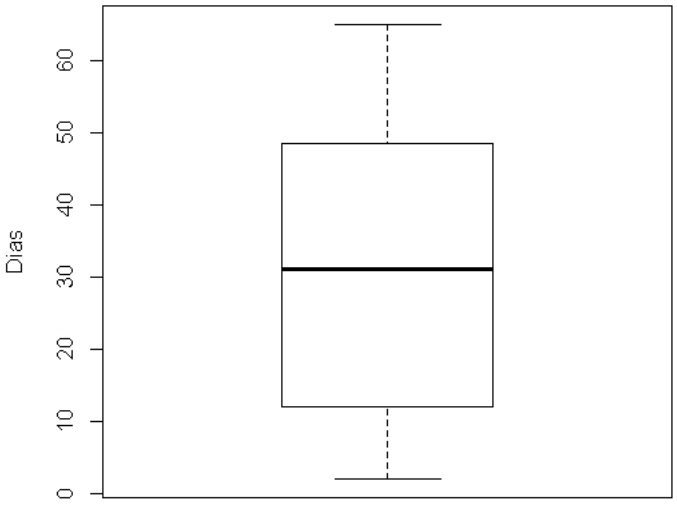


Figura 3.3: Boxplot do tempo de internação.



(3.3) e seus respectivos erros padrão.

Pela Tabela 3.5, pode-se notar que não há uma grande diferença das estimativas, entre os modelos, dos parâmetros α e β.

Considere a medida do quadrado médio de previsão, expressa por

Figura 3.4: Gráfico de dispersão do logaritmo do índice de prognóstico para recuperação a longo prazo versus tempo de internação.

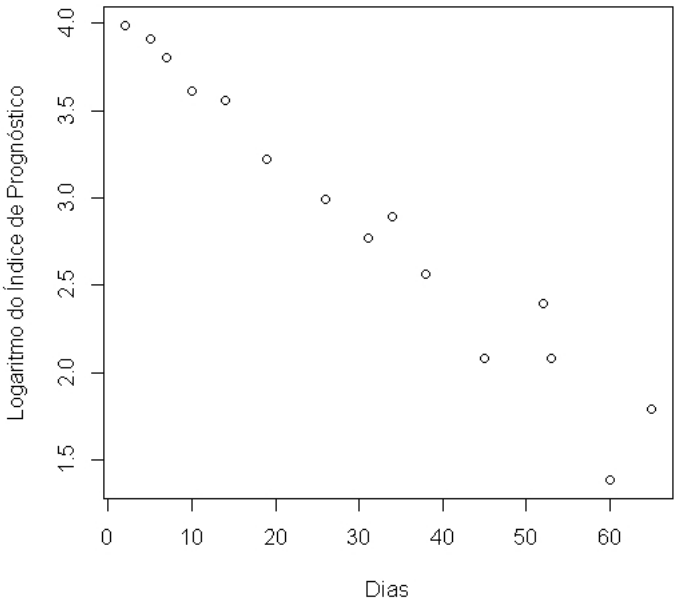


Tabela 3.5: Estimativas dos parâmetros (erro padrão aproximado) para o modelo em (3.3) ajustado aos dados logaritmo do índice de prognóstico.

Modelo	Normal		PR-EM	
Parâmetro	Estimativa	Erro Padrão	Estimativa	Erro Padrão
α	4,0372	(0,0841)	4,0223	(0,1304)
β	-0,0380	(0,0023)	-0,0370	(0,0027)

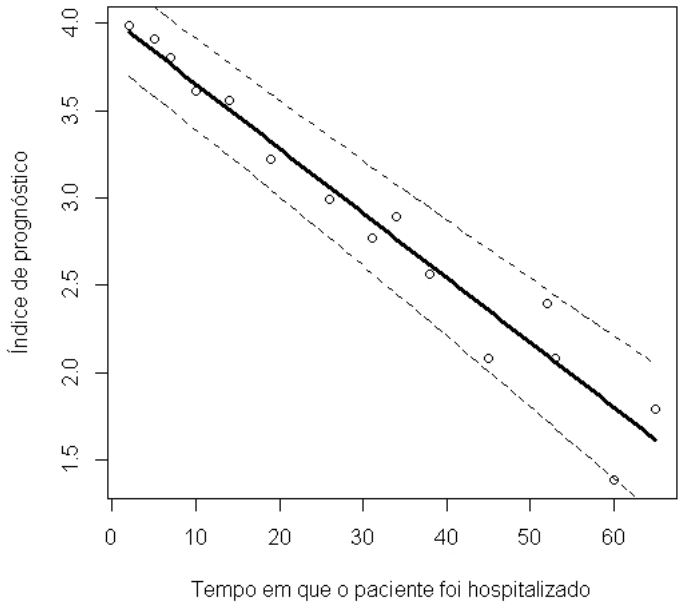
$$QMP = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 .$$

As estimativas de QMP para o modelos sob erros Normal e através do algoritmo PR-EM são 0,0279 e 0,0285, respectivamente. Observa-se que o modelo ajustado por meio do algoritmo PR-EM possui um erro de previsão maior que o modelo ajustado sob erros normais, porém essas estimativas tem uma diferença de 0,0006.

A Figura 3.5 apresenta a reta ajustada do modelo obtida por meio do algoritmo PR-EM e seu intervalo de confiança com nível de 95%. Pode-se notar que a curva aparenta ajustar a relação entre índice de prognóstico e o tempo em que o paciente foi hospitalizado.

Desta forma, o algoritmo PR-EM parece ser uma boa alternativa na análise desses dados, uma vez que, obteve uma boa representação da relação entre o índice de prognóstico e o tempo em que o paciente foi hospitalizado. Além disso, não houve suposição alguma sobre a verdadeira distribuição da variável resposta.

Figura 3.5: Gráfico de dispersão do índice de prognóstico versus tempo de internação do paciente e sua reta ajustada do modelo com seu intervalo de confiança com nível de 95%.

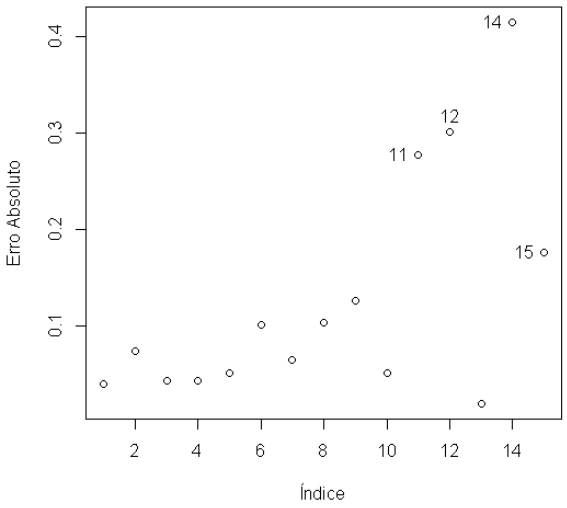


Defini-se por erro de estimação absoluto a medida

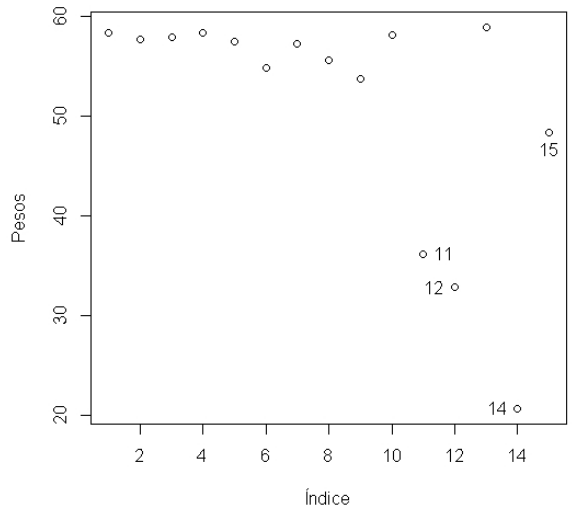
$$|y_i - \hat{y}_i|.$$

Pelo gráfico de dispersão do erro de estimação absoluto versus os índices (Figura 3.7(a)) destaca-se as observações 11, 12, 14 e 15 como as observações com o maior erro de estimação que são as mesmas destacadas como as observações com menor peso (Figura 3.7(b)) na estimação dos parâmetros. Desta forma, o algoritmo PR-EM tende a penalizar observações que indiquem poder exercer alguma influência na estimativa dos seus parâmetros.

Figura 3.6: Erro de estimação absoluto e os pesos para o modelo em (3.3) obtido através do algoritmo PR-EM.



(a) Erro de estimação absoluto



(b) Pesos

CAPÍTULO 4

Considerações finais e Trabalhos Futuros

Nesta dissertação foram estudados modelos de regressão linear e não linear com erros sob distribuição de mistura de escala normal (MRL-MEN e MRNL-MEN). Foi proposto um estimador para os parâmetros dos MRNL-MEN no caso em que não há a especificação da distribuição da variável de mistura e o parâmetro de escala é conhecido e igual a 1. Outra proposta, foram duas metodologias para a estimação da matriz de variância e covariância desses estimadores, que pode ser utilizado, também, para o caso linear.

Estudos de simulação foram feitos para avaliar o comportamento dos estimadores propostos e a robustez desses mesmos estimadores na presença de observações atípicas. Baseado nos resultados, observa-se que o algoritmo PR-EM produz estimadores com boas propriedades assintóticas para o caso dos MRNL-MEN. Pode-se observar também que os estimadores obtidos por meio do algoritmo PR-EM possuem uma robustez em suas estimativas sendo que, estes estimadores, produzem mudanças relativas menores que a do MRNL-N. Como forma de ilustrar as metodologias apresentadas no decorrer desta dissertação, fez-se uma aplicação das mesmas em um conjunto de dados reais. Pode-se observar que as estimativas obtidas por meio do algoritmo PR-EM apresentou ajuste similar a de outros modelos estudados. Além disso, ao observar a curva ajustada e seu intervalo de confiança de 95% apresentado na Figura , temos que o ajuste através do algoritmo PR-EM apresenta um bom ajuste da curvatura do modelo.

Como propostas para trabalhos futuros pode se destacar o desenvolvimento da análise de diagnóstico e influência para modelos de regressão linear e não linear sob erros com distribuição de mistura de escala normal sem especificação da variável de mistura com parâmetro de escala fixo e igual a um.

Referências

ANDREWS, D.; MALLOWS, S. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, Wiley for the Royal Statistical Society, v. 36, n. 1, p. 99–102, 1974.

ARELLANO-VALLE, R. *Distribuições elípiticas: propriedades, inferência e aplicações a modelos de regressão*. Tese (Doutorado) — Instituto de Matemática e Estatística - Universidade de São Paulo, 1994.

BASSO, R. M.; LACHOS, V. H.; CABRAL, C. R. B.; GHOSH, P. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 12, p. 2926–2941, 2010.

BOX, G.; TIAO, G. *Bayesian Inference in Statistical Analysis*. Nova Jersey: Reading, Mass: Addison-Wisley, 1973.

COX, D. R.; HINKLEY, D. V. *Theoretical statistics*. Flórida: CRC Press, 1979.

DEMPSTER, A.; LAIR, N.; RUBIN, D. Iteratively reweighted least squares for linear regression when errors are normal/independent distributed. *In Multivariate Analysis*, North-Holland, Amsterdam, V, p. 35–57, 1980.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, JSTOR, p. 1–38, 1977.

EFRON, B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, Institute of Mathematical Statistics, Bethesda, v. 7, p. 1–26, 1979.

EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. Flórida: CRC press, 1994.

EGUCHI, S.; COPAS, J. Interpreting kullback–leibler divergence with the neyman–pearson lemma. *Journal of Multivariate Analysis*, Elsevier, v. 97, n. 9, p. 2034–2040, 2006.

FANG, K.; KOTZ, S.; NG, K. *Symmetric Multivariate and Related Distributions*. Londres: Chapman and Hall, 1990.

GARAY, A. M.; LACHOS, V. H.; BOLFARINE, H.; CABRAL, C. R. Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*, Springer, p. 1–32, 2015.

GARAY, A. M.; LACHOS, V. H.; LIN, T.-I. Nonlinear censored regression models with heavy-tailed distributions. *STATISTICS AND ITS INTERFACE*, v. 9, n. 3, p. 281–293, 2016.

GHOSH, J.; TOKDAR, S. Convergence and consistency of newton's algorithm for estimating mixing distributions. In: FAN, J.; KOUL, H. (Ed.). *Frontiers in Statistics*. London: Imp. Coll. Press, 2006. p. 429–443.

KUTNER, M.; NACHTSHEIM, C.; NETER, J.; LI, W. *Applied linear statistical models*. Nova York: McGraw Hill, 2004.

LIN, T. I. Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, v. 100, n. 2, p. 257–265, 2009.

LOUIS, T. Finding the observed information matrix when using the em algorithm. *Jornal of the Royal Statistic Society*, Wiley for the Royal Statistical Society, Hoboken, v. 44, n. 2, p. 226–233, 1982.

MARTIN, R.; GHOSH, J. Stochastic approximation and newton's estimate of a mixing distribution. *Statistic Science*, v. 23, p. 365–382, 2008.

MARTIN, R.; HAN, Z. A semiparametric scale-mixture regression model and predictive recursion maximum likelihood. *Computational Statistic and Data Analysis*, Elsevier, Amsterdã, v. 94, p. 75–85, 2016.

MARTIN, R.; TOKDAR, S. Semiparametric inference in mixture models with predictive recursion marginal likelihood. *Biometrika*, Oxford Journals, Oxford, v. 98, n. 3, p. 567–582, 2011.

MARTIN, R.; TOKDAR, S. A nonparametric empirical bayes framework for large-scale multiple testing. *Biostatistic*, Oxford Journals, Oxford, v. 13, p. 427–439, 2012.

MCLACHLAN, G.; KRISHNAN, T. *The EM algorithm and extensions*. Nova Jersey: John Wiley & Sons, 2007. v. 382.

MEILIJSON, I. A fast improvement to the em algorithm on its own terms. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 127–138, 1989.

MEZA, C.; OSORIO, F.; CRUZ, R. la. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Stat Comput*, v. 22, p. 121–139, 2012.

MOORE, D.; MCCABE, G.; DUCKWORTH, W.; SCLOVE, S. *The Practice of Business Statistics: Using Data for Decisions*. 2. ed. Macmillan: W. H. Freeman, 2006. 859 p.

NEWTON, M.; QUINTANA, F.; ZHANG, Y. Nonparametric bayes methods using predictive updating. *Practical Nonparametric and Semiparametric Bayesian Statistics. In: Lecture Notes in Statistics*, Springer, Nova York, v. 133, p. 45–61, 1998.

NEWTON, M. A. On a nonparametric recursive estimator of the mixing distribution. *Sankhyā: The Indian Journal of Statistics, Series A*, JSTOR, p. 306–322, 2002.

TOKDAR, S.; MARTIN, R.; GHOSH, J. Consistency of a recursive estimate of mixing distributions. *The Annals of Statistics*, Institute of Mathematical Statistics, Ohio, v. 37, n. 5A, p. 2502–2522, 2009.

WANG, W.-L.; LIN, T.-I. Robust model-based clustering via mixtures of skew-t distributions with missing information. *Advances in Data Analysis and Classification*, Springer, v. 9, n. 4, p. 423–445, 2015.