

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

MODELOS MISTOS SEMIPARAMÉTRICOS  
PARCIALMENTE NÃO LINEARES

Robson José Mariano Machado

São Carlos  
2014

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

MODELOS MISTOS SEMIPARAMÉTRICOS  
PARCIALMENTE NÃO LINEARES

Robson José Mariano Machado

Dissertação apresentada ao Programa de Pós-graduação em Estatística como requisito à obtenção do título de Mestre em Estatística.

Orientadora: Dra. Cibele Maria Russo  
Agência financiadora: Capes

São Carlos  
2014

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

M149mm Machado, Robson José Mariano.  
Modelos mistos semiparamétricos parcialmente não lineares / Robson José Mariano Machado. -- São Carlos : UFSCar, 2014.  
52 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2014.

1. Análise de regressão. 2. Modelos semiparamétricos. 3. Modelos mistos não-lineares. 4. Diagnóstico de influência local. I. Título.

CDD: 519.536 (20ª)



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**  
Centro de Ciências Exatas e de Tecnologia  
Programa de Pós-Graduação em Estatística  
Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40  
FONE: (016) 3351-8292 – Email: ppgest@ufscar.br  
13565-905 - SÃO CARLOS-SP - BRASIL

---

## FOLHA DE APROVAÇÃO

**Aluno(a) : Robson José Mariano Machado**

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 28/03/2014  
PELA COMISSÃO JULGADORA:

Presidente

Profa. Dra. Cibele Maria Russo Noveli (ICMC-USP/Orientadora)

1º Examinador

Prof. Dr. Adriano Polpo de Campos (DEs-UFSCar)

2º Examinador

Prof. Dr. Cristian Marcelo Villegas Lobos (ESALQ-USP)

# Resumo

Dados correlacionados com estrutura não linear são comuns em bioestatística, estudos farmacocinéticos e longitudinais. Modelos mistos não lineares são ferramentas úteis para se analisar esses tipos de problemas. Nesta dissertação, propõe-se uma generalização desses modelos, chamada de modelo misto semiparamétrico parcialmente não linear (MMSPNL), com uma função não paramétrica para se modelar a média da variável resposta. Assume-se que a média da variável de interesse é explicada por uma função não linear, que depende de parâmetros de efeitos fixos e variáveis explicativas, e por uma função não paramétrica. Tal função não paramétrica possui grande flexibilidade, permitindo uma melhor adequação à forma funcional que subjaz aos dados. Os efeitos aleatórios são incluídos linearmente ao modelo, o que simplifica a expressão da distribuição da variável resposta e permite considerar a estrutura de correlação intra grupo. É assumido que os erros aleatórios e efeitos aleatórios conjuntamente seguem uma distribuição normal multivariada. Em relação a função não paramétrica, utiliza-se a técnica de suavização com *P-splines*. A metodologia para se obterem as estimativas dos parâmetros é o método de máxima verossimilhança penalizada. Os efeitos aleatórios podem ser obtidos usando-se o método de Bayes empírico. A qualidade do modelo e a identificação de observações aberrantes é verificada pelo método de influência local e por análise de resíduos. O conjunto de dados farmacocinéticos, em que o antiasmático *theophylline* foi administrado a 12 sujeitos e concentrações séricas foram tomadas em 11 instantes de tempo durante as 25 horas (após ser administrado), foi reanalisado com o modelo proposto, exemplificando seu uso e propriedades.

**Palavras-chave:** modelos mistos não lineares; modelos semiparamétricos; suavização; influência local; *P-splines*.

# Abstract

Correlated data sets with nonlinear structure are common in many areas such as biostatistics, pharmacokinetics and longitudinal studies. Nonlinear mixed-effects models are useful tools to analyse those type of problems. In this dissertation, a generalization to this models is proposed, namely by semiparametric partially nonlinear mixed-effects model (MMSPNL), with a nonparametric function to model the mean of the response variable. It assumes that the mean of the interest variable is explained by a nonlinear function, which depends on fixed effects parameters and explanatory variables, and by a nonparametric function. Such nonparametric function is quite flexible, allowing a better adequacy to the functional form that underlies the data. The random effects are included linearly to the model, which simplify the expression of the response variable distribution and enables the model to take into account the within-group correlation structure. It is assumed that the random errors and the random effects jointly follow a multivariate normal distribution. Relate to the nonparametric function, it is deal with the P-splines smoothing technique. The methodology to obtain the parameters estimates is penalized maximum likelihood method. The random effects may be obtained by using the Empirical Bayes method. The goodness of the model and identification of potencial influent observation is verified with the local influence method and a residual analysis. The pharmacokinetic data set, in which the anti-asthmatic drug theophylline was administered to 12 subjects and serum concentrations were taken at 11 time points over the 25 hours (after being administered), was re-analysed with the proposed model, exemplifying its uses and properties.

**Key words:** nonlinear mixed-effects models; semiparametric models; smoothing; local influence; P-splines.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Modelos não lineares com efeitos mistos</b>	<b>4</b>
2.1	Introdução . . . . .	4
2.2	Especificação do modelo . . . . .	5
2.3	Estimação de máxima verossimilhança . . . . .	7
2.4	Predição dos efeitos aleatórios . . . . .	8
<b>3</b>	<b>Modelo mistos semiparamétrico parcialmente não linear</b>	<b>10</b>
3.1	Introdução . . . . .	10
3.2	Especificação do modelo . . . . .	11
3.3	Representação matricial . . . . .	14
3.4	Pressupostos do modelo . . . . .	14
3.5	Função de log-verossimilhança penalizada . . . . .	15
3.6	Processo de estimação . . . . .	17
3.6.1	O parâmetro de suavização . . . . .	18
3.6.2	Predição dos efeitos aleatórios e da variável resposta . . . . .	19
3.6.3	Erros padrão das estimativas . . . . .	19
<b>4</b>	<b>Influência local</b>	<b>20</b>
4.1	Introdução . . . . .	20
4.2	Método de influência local . . . . .	21
4.3	Derivação da curvatura . . . . .	24
4.3.1	Matriz de informação . . . . .	24
4.4	Matriz de perturbação . . . . .	26
4.4.1	Ponderação de casos . . . . .	26
4.4.2	Perturbação na matriz escala . . . . .	27
4.4.3	Perturbação na variável resposta . . . . .	28
<b>5</b>	<b>Aplicação</b>	<b>29</b>
5.1	Dados de concentração de <i>theophylline</i> . . . . .	29
5.2	Modelo proposto . . . . .	30

5.3	Análise de resíduos . . . . .	33
5.4	Diagnóstico de Influência . . . . .	35
<b>6</b>	<b>Discussão e propostas futuras</b>	<b>41</b>
<b>A</b>	<b>Dados de concentração de theophylline</b>	<b>46</b>
<b>B</b>	<b><i>B</i>-splines e <i>P</i>-splines</b>	<b>50</b>



# Lista de Figuras

5.1	Diagrama de dispersão dos dados de concentração de <i>theophylline</i> . . . . .	29
5.2	Boxplot dos dados de concentração de <i>theophylline</i> . . . . .	30
5.3	Modelos paramétrico e semiparamétrico ajustados . . . . .	32
5.4	Gráfico de índice dos efeitos aleatórios preditos para o modelo semiparamétrico . . . . .	33
5.5	Distância de Mahalanobis para os modelos paramétrico e semiparamétrico	34
5.6	Resíduos condicionais padronizados para os modelos paramétrico e semiparamétrico . . . . .	34
5.7	Gráficos de quantis dos resíduos condicionais padronizados. . . . .	35
5.8	Autovalores relativos às direções de influência sob o esquema de ponderação de casos . . . . .	36
5.9	Gráficos de índices de $m(0)$ para $\hat{\theta}$ sob o esquema de ponderação de casos.	36
5.10	Gráficos de índices de $m(2)$ para $\hat{\theta}$ sob o esquema de ponderação de casos.	37
5.11	Autovalores relativos às direções de influência sob o esquema de perturbação na matriz escala. . . . .	37
5.12	Gráficos de índices de $m(0)$ para $\hat{\theta}$ sob o esquema perturbação na matriz escala. . . . .	38
5.13	Gráficos de índices de $m(2)$ para $\hat{\theta}$ sob o esquema de perturbação na matriz escala. . . . .	38
5.14	Autovalores relativos às direções de influência sob o esquema de perturbação na variável resposta . . . . .	39
5.15	Gráficos de índices de $m(0)$ para $\hat{\theta}$ sob o esquema de perturbação na variável resposta. . . . .	39
5.16	Gráficos de índices de $m(6)$ para $\hat{\theta}$ sob o esquema de perturbação na variável resposta. . . . .	40
B.1	Funções base de <i>B-splines</i> de terceiro grau e 3 nós interiores. . . . .	51

# Lista de Tabelas

5.1	Estimadores e seus respectivos erros padrão para os modelos paramétrico e semiparamétrico. . . . .	32
A.1	Dados de concentração de theophylline . . . . .	46

# Capítulo 1

## Introdução

Modelos não lineares são ferramentas úteis para se analisarem dados que apresentam relação não linear entre as covariáveis e as variáveis respostas, ao considerarem expressões matemáticas para explicar o mecanismo gerador dos dados. Como vantagem desses modelos tem-se a interpretação dos parâmetros e bons resultados na predição da variável resposta fora do domínio dos dados observados (Pinheiro & Bates 2000). Por outro lado, há que se considerar também que esses modelos apresentam dificuldades encontradas, por exemplo, no processo de estimação, uma vez que em geral as estimativas dos parâmetros não possuem solução analítica e tornam-se necessárias aproximações numéricas.

Dados com medidas repetidas, caracterizados pela correlação entre observações intraunidade experimental, também podem apresentar relação não linear entre a variável resposta e os parâmetros. Para modelar a correlação intraunidade experimental uma possibilidade é considerar modelos com efeitos aleatórios. Davidian & Giltinan (1995) apresentam o modelo não linear com efeitos aleatórios, que são úteis para se estudarem tais dados. Nesse caso, a variável resposta é explicada por uma função não linear nos parâmetros e nos efeitos aleatórios. A dificuldade desses modelos é a necessidade de integrações numéricas para se obter a distribuição da variável resposta (modelo marginal). Uma alternativa é considerar o modelo misto parcialmente não linear (Vonesh & Carter 1992) em que os efeitos aleatórios são incluídos linearmente. Neste trabalho assume-se que os erros e efeitos aleatórios seguem uma distribuição normal. A vantagem dessa formulação surge ao passo em que o modelo marginal é obtido de forma direta por propriedades da distribuição normal. Porém, os efeitos aleatórios não possuem a mesma interpretação dos parâmetros. Russo (2010) estende esse modelo considerando distribuições elípticas, que fornecem estimativas robustas para os parâmetros e podem acomodar observações aberrantes.

Ke & Wang (2001) apresentam um modelo semiparamétrico misto não linear em que a variável resposta é explicada por uma função que pode ser não linear com relação a parâmetros, função não paramétrica e efeitos aleatórios. Assim, torna-se necessário

utilizar aproximações numéricas para se obter o logaritmo da função de verossimilhança para o modelo marginal. A complexidade desse modelo dificulta o processo de estimação e inferência, tanto dos parâmetros de efeitos fixos quanto da função não paramétrica. Elmi et al. (2011) desenvolveram um método alternativo para ajustar o modelo apresentado por Ke & Wang (2001), modelando a função não paramétrica por meio de *B*-splines. Esse fato fornece vantagens computacionais na estimação dos parâmetros. Ibacache-Pulgar et al. (2012) desenvolveram os modelos mistos lineares semiparamétricos aditivos considerando distribuições elípticas para as componentes aleatórias.

Desta forma, propõe-se o modelo misto semiparamétrico parcialmente não linear que pode ser visto como generalização direta do modelo proposto por Vonesh & Carter (1992), ao incluir uma função não paramétrica para modelar a média da variável resposta. Os parâmetros fornecem um resumo interpretável dos dados e a função não paramétrica fornece flexibilidade com relação à forma funcional que subjaz aos dados.

Além disso, desenvolve-se o método de influência local, proposto por Cook (1986), para o modelo misto semiparamétrico parcialmente não linear. Em particular, será considerada a curvatura normal conformal (Poon & Poon 1999), ao invés da curvatura normal, que é invariante sob reparametrizações conformais e é uma medida padronizada. Essa técnica de diagnóstico propõe um enfoque geométrico para se verificarem as variações que pequenas perturbações nos dados ou nos pressupostos do modelo podem causar nas estimativas dos parâmetros. Alguns trabalhos relacionados a esse método na literatura são citados a seguir. Russo et al. (2009) realizaram estudo de influência local para o modelo misto parcialmente não linear com distribuições elípticas. Osório et al. (2007) desenvolveram o método de influência local para modelos mistos lineares aplicados a dados com estrutura longitudinal. Ibacache-Pulgar et al. (2012) realizaram estudos de influência local para o modelo misto semiparamétrico aditivo considerando distribuições elípticas para as componentes aleatórias.

A organização desta dissertação está descrita a seguir. Após esta introdução, no Capítulo 2 é apresentada uma visão geral da teoria de modelos não lineares mistos paramétricos, destacando-se as vantagens e desvantagens. No Capítulo 3 é apresentado o modelo misto semiparamétrico parcialmente não linear e são realizadas discussões importantes sobre os métodos de estimação e inferências. No Capítulo 4 é desenvolvido o método de influência local para o modelo proposto. No Capítulo 5 é apresentada a análise do conjunto de dados farmacocinéticos *theophylline*, em que o antiasmático foi administrado a 12 sujeitos e 11 medidas dessa substância foram tomadas nas 25 horas seguintes. No Capítulo 6 tem-se a discussão do estudo realizado e indicação de propostas futuras.

# Capítulo 2

## Modelos não lineares com efeitos mistos

Neste capítulo é apresentada uma visão geral da teoria de modelos não lineares com efeitos aleatórios. Na Seção 2.1 são discutidas possíveis aplicações e alguns trabalhos relacionados a essa classe de modelos. Na Seção 2.2 é definido o modelo não linear com efeitos mistos, de duas formas distintas, ressaltando-se as vantagens e desvantagens de se trabalhar com cada modelo. Na Seção 2.3 são apresentadas as equações obtidas de estimação baseadas no logaritmo da função de verossimilhança e é apontado o algoritmo de Fisher para se obterem as estimativas de máxima verossimilhança dos parâmetros associados ao modelo apresentado. Na Seção 2.4 é apresentado o método de Bayes empírico para predição dos efeitos aleatórios.

### 2.1 Introdução

Dados com medidas repetidas, comuns em bioestatística, tais como dados farmacocinéticos e dados de crescimento, são, em geral caracterizados por uma relação não linear entre as covariáveis e a variável resposta. Podem-se considerar modelos de regressão polinomial que apresentam bons ajustes aos dados. Porém, muitas vezes estes não permitem interpretação dos parâmetros e podem apresentar problemas para predição da variável resposta para valores do domínio dos dados além do intervalo (ou região) a que pertencem os dados observados. Uma alternativa são os modelos não lineares que visam descrever a mecânica geradora da resposta e, portanto, possuem parâmetros interpretáveis. Além disso, esses modelos possuem boa capacidade de predição para valores do domínio dos dados além dos observados, já que levam em consideração o mecanismo gerador dos mesmos. Dados com medidas repetidas requerem modelos que levem em consideração a estrutura de correlação existente nas observações intraunidade experimental. Modelos com efeitos mistos desempenham bem essa função ao adiciona-

rem efeitos aleatórios ao modelo. Dessa forma, modelos não lineares com efeitos mistos são úteis para se analisarem esses dados, uma vez que modelam tanto seu mecanismo quanto a estrutura de correlação intraunidade experimental.

Existem diversas maneiras de se formularem esses modelos. Lindstrom & Bates (1990) e Pinheiro & Bates (2000) consideram o caso em que os parâmetros de efeitos fixos e efeitos aleatórios são incorporados de forma não linear. A dificuldade em utilizar essa formulação vem do fato de que as inferências são baseadas no modelo marginal que, em geral, não possui solução analítica. Dessa forma, é necessário realizar aproximações numéricas para o logaritmo da função de verossimilhança, o que resulta em perda da capacidade de predição do modelo. Uma alternativa é considerar o modelo apresentado em Vonesh & Carter (1992), em que a variável resposta é explicada por uma função não linear nos parâmetros fixos e uma forma linear nos efeitos aleatórios. Nesse caso, a distribuição marginal da variável resposta é conhecida e as inferências podem ser realizadas de maneira direta. Neste capítulo será considerado principalmente o último modelo, considerando que o vetor de erros é independente do vetor de efeitos aleatórios e que ambos seguem uma distribuição normal multivariada.

## 2.2 Especificação do modelo

Seja  $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^\top$  o vetor de observações do  $i$ -ésimo indivíduo ( $i = 1, \dots, n$ ). O modelo não linear usual com efeitos mistos é representado por

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\eta}_i(\boldsymbol{\phi}_i, \mathbf{x}_i) + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{\phi}_i &= \mathbf{A}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i, \end{aligned} \tag{2.1}$$

em que  $\boldsymbol{\eta}_i(\boldsymbol{\phi}_i, \mathbf{x}_i) = (\eta_i(\phi_{i1}, \mathbf{x}_{i1}), \dots, \eta_i(\phi_{im_i}, \mathbf{x}_{im_i}))^\top$  é uma função não linear do vetor de parâmetros  $\boldsymbol{\phi}_i$ , com  $\boldsymbol{\phi}_i = (\phi_{i1}, \dots, \phi_{im_i})^\top$  que depende das matrizes de planejamento  $\mathbf{A}_i = (\mathbf{A}_{i1}, \dots, \mathbf{A}_{im_i})^\top$  e  $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{im_i})^\top$  com dimensões  $m_i \times p$  e  $m_i \times q$ , respectivamente,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de parâmetros referentes aos efeitos fixos,  $\mathbf{u}_i$  é um vetor  $q$ -dimensional de efeitos aleatórios,  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})$  é uma matriz de covariáveis e  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  é o vetor de erros aleatórios.

Denote por  $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  a distribuição normal multivariada com vetor de médias  $\boldsymbol{\mu}$   $n$ -dimensional e matriz de variâncias-covariâncias  $\boldsymbol{\Sigma}_{n \times n}$ . É comum supor que

$$\mathbf{u}_i \sim N_q(\mathbf{0}, \mathbf{Q}) \quad \text{e} \quad \boldsymbol{\varepsilon}_i \sim N_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i})$$

são mutuamente independentes para  $i = 1, \dots, n$ , em que  $\mathbf{Q}$  é a matriz variâncias-covariâncias de  $\mathbf{u}_i$  e  $\mathbf{I}_{m_i}$  representa a matriz identidade de dimensão  $m_i$ .

As estimativas de máxima verossimilhança do modelo (2.1) são baseadas na distri-

buição de probabilidades de  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top$ , que é fornecida por

$$\begin{aligned} \mathbf{p}(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{Q}) &= \int \mathbf{p}(\mathbf{y}, \mathbf{u}|\boldsymbol{\beta}, \sigma^2, \mathbf{Q})d\mathbf{u} \\ &= \int \mathbf{p}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma^2)\mathbf{p}(\mathbf{u}|\mathbf{Q})d\mathbf{u}, \end{aligned} \quad (2.2)$$

em que  $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_q)^\top$ ,  $\mathbf{p}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma^2)$  é a função densidade condicional de  $\mathbf{y}$  dados os efeitos aleatórios  $\mathbf{u}$  e  $\mathbf{p}(\mathbf{u}|\mathbf{Q})$  representa a função de densidade de  $\mathbf{u}$ . É conveniente expressar a matriz de variâncias-covariâncias dos efeitos aleatórios na forma de um *fator de precisão relativa*<sup>1</sup>,  $\boldsymbol{\Delta}$ , tal que  $\mathbf{Q}^{-1} = \sigma^2\boldsymbol{\Delta}^\top\boldsymbol{\Delta}$  (Pinheiro & Bates 2000, cap. 7).

Assim, é possível expressar a função densidade de probabilidades de  $\mathbf{y}$  como

$$\mathbf{p}(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}) = \frac{|\boldsymbol{\Delta}|^n}{(2\pi\sigma^2)^{(m+nq)/2}} \prod_{i=1}^n \int \exp \left\{ -\frac{\|\mathbf{y}_i - \boldsymbol{\eta}_i(\boldsymbol{\beta}, \mathbf{u}_i)\|^2 + \|\boldsymbol{\Delta}\mathbf{u}_i\|^2}{2\sigma^2} \right\} d\mathbf{u}_i \quad (2.3)$$

em que  $\boldsymbol{\eta}_i(\boldsymbol{\beta}, \mathbf{u}_i) = \boldsymbol{\eta}_i(\boldsymbol{\phi}(\boldsymbol{\beta}, \mathbf{u}_i), \mathbf{x}_i)$ .

Como  $\boldsymbol{\eta}_i$  pode ser não linear nos efeitos aleatórios, a integral em (2.3) pode não ter solução analítica. Para facilitar a maximização da função de verossimilhança, diferentes aproximações dessa equação têm sido propostas (vide, por exemplo, Pinheiro & Bates 2000, pag. 313). Um desses métodos, considerado por Vonesh & Carter (1992), consiste em realizar uma expansão de Taylor de primeira ordem da função  $\boldsymbol{\eta}_i(\boldsymbol{\phi}_i, \mathbf{x}_i)$  em torno de  $\mathbf{u}_i = 0$ . Levando-se em conta que para cada  $i \in \{1, \dots, n\}$  a forma funcional de  $\boldsymbol{\eta}_i$  é a mesma, faz sentido considerar  $\boldsymbol{\eta}_i(\boldsymbol{\phi}_i, \mathbf{x}_i) = \boldsymbol{\eta}(\boldsymbol{\phi}_i, \mathbf{x}_i)$ . Assim, essa aproximação do modelo (2.1) é fornecida por

$$\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{Z}_i^* \mathbf{u}_i^* + \boldsymbol{\varepsilon}_i, \quad (2.4)$$

em que

$$\mathbf{Z}_i^* = \left[ \frac{\partial \boldsymbol{\eta}(\boldsymbol{\phi}_i, \mathbf{x}_i)}{\partial \boldsymbol{\phi}_i^\top} \Big|_{\boldsymbol{\phi}_i = \mathbf{A}_i \boldsymbol{\beta}} \right] \mathbf{A}_i$$

e  $\mathbf{u}_i^* = \mathbf{u}_i$ , denominado modelo pseudo não linear com efeitos mistos.

Esse modelo pode ser visto como um caso particular do proposto por Vonesh & Carter (1992) denominado modelo de regressão não linear com efeitos mistos generalizado

$$\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (2.5)$$

com  $\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) = (\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_{i1}), \dots, \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_{im_i}))^\top$  uma função não linear de  $\boldsymbol{\beta}$  com dimensão  $m_i \times 1$ ,  $\boldsymbol{\beta}$  vetor de parâmetros fixos,  $\mathbf{x}_i$  vetor de variáveis explicativas,  $\mathbf{Z}_i$  matriz de planejamento<sup>2</sup>,  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  e  $\mathbf{u}_i = (u_{i1}, \dots, u_{iq})^\top$  são os vetores de erros e efeitos

<sup>1</sup>Se  $\mathbf{Q}$  for positiva definida, então  $\boldsymbol{\Delta}$  existe, mas não é única.

<sup>2</sup>Exemplos de estruturas para a matriz  $\mathbf{Z}_i$  são apontados na página 14.

aleatórios, respectivamente, mutuamente independentes. Neste trabalho o modelo (2.5) será referido como modelo misto parcialmente não linear. Vonesh & Carter (1992) consideram o modelo (2.5) supondo que os erros e efeitos aleatórios são independentes e seguem uma distribuição normal. Russo (2010) generaliza o último modelo supondo distribuições elípticas para as componentes aleatórias, que possibilitam ao modelo, maior capacidade de acomodar observações aberrantes. Neste trabalho será dada ênfase ao modelo (2.5) em que os erros e efeitos aleatórios são mutuamente independentes e seguem uma distribuição normal, o que equivale a supor que

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{u}_i \end{pmatrix} \stackrel{ind.}{\sim} N_{m_i+q} \left( \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \mathbf{Z}_i \mathbf{Q} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i} & \mathbf{Z}_i \mathbf{Q} \\ \mathbf{Q} \mathbf{Z}_i^\top & \mathbf{Q} \end{bmatrix} \right), \quad (2.6)$$

$i = 1, \dots, n$ . Como vantagem da formulação (2.5) e (2.6) tem-se que a distribuição de probabilidades de  $\mathbf{y}$  é obtida de maneira direta pelas propriedades da distribuição normal multivariada, o que facilita a maximização da função de verossimilhança. Pela mesma razão, a distribuição de probabilidades dos efeitos aleatórios  $\mathbf{u}_i$  dado  $\mathbf{y}_i$  é facilmente determinada, o que permite prever os efeitos aleatórios pelo método Bayes empírico. Além disso, tal formulação preserva a média da distribuição da variável resposta em sua forma original não linear e possibilita a inclusão da correlação intragrupo.

## 2.3 Estimação de máxima verossimilhança

A componente da função de log-verossimilhança  $\sum_{i=1}^n L_i(\mathbf{y}_i; \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$  para o modelo (2.6), associada ao grupo  $i$  é fornecida por

$$L_i = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \frac{1}{2} [\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i)]^\top \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i)] \quad (2.7)$$

para  $i = 1, \dots, n$ , em que  $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{Q} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i}$ . O vetor de parâmetros a ser estimado é definido como  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\tau}^\top)^\top$ , em que  $\boldsymbol{\tau}$  é o vetor com elementos da matriz  $\mathbf{Q}$ . Defina  $\gamma_0 = \sigma^2$  e  $\gamma_j = \tau_j$ ,  $j = 1, \dots, t$ . Os elementos da função escore (Russo et al. 2009) para  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma_0, \dots, \gamma_t)^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$  são dados por

$$\begin{aligned} \mathbf{U}_\beta &= \sum_{i=1}^n \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \text{ e} \\ \mathbf{U}_\gamma &= (U_{\gamma_0}, \dots, U_{\gamma_t})^\top, \text{ com} \\ U_{\gamma_j} &= -\frac{1}{2} \sum_{i=1}^n \left\{ \text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \right] - \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i^{-1}(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \right\}, \end{aligned}$$



em que  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i))$ ,  $\mathbf{J}_i = \partial \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) / \partial \boldsymbol{\alpha}^\top$  e  $\dot{\Sigma}_i(j) = \partial \Sigma_i / \partial \gamma_j$ ,  $j = 0, \dots, t$ ,  $i = 1, \dots, n$ .

A matriz de informação de Fisher para  $\boldsymbol{\theta}$  é dada por

$$\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\gamma\gamma} \end{pmatrix}, \quad (2.8)$$

em que

$$\begin{aligned} \mathbf{K}_{\beta\beta} &= \sum_{i=1}^n \mathbf{J}_i^\top \Sigma_i^{-1} \mathbf{J}_i \\ \mathbf{K}_{\gamma\gamma} &= [\mathbf{K}_{\gamma_r, \gamma_s}], \text{ cujo } (r, s)\text{-ésimo elemento é dado por} \\ \mathbf{K}_{\gamma_r, \gamma_s} &= \frac{1}{2} \sum_{i=1}^n \text{tr}[\Sigma_i^{-1} \dot{\Sigma}_i(r) \Sigma_i^{-1} \dot{\Sigma}_i(s)]. \end{aligned}$$

As estimativas de máxima verossimilhança para  $\boldsymbol{\theta}$  podem ser obtidas com o algoritmo escore de Fisher, que é dado por

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{(k+1)} &= \hat{\boldsymbol{\beta}}_{(k)} + [\mathbf{K}_{\beta\beta}]_{(k)}^{-1} \mathbf{U}_{\beta(k)} \\ \hat{\boldsymbol{\gamma}}_{(k+1)} &= \hat{\boldsymbol{\gamma}}_{(k)} + [\mathbf{K}_{\gamma\gamma}]_{(k)}^{-1} \mathbf{U}_{\gamma(k)} \end{aligned}$$

para  $k = 0, 1, 2, \dots$ . As estimativas iniciais  $\hat{\boldsymbol{\alpha}}_{(0)}$  e  $\hat{\boldsymbol{\gamma}}_{(0)}$  podem ser as estimativas de mínimos quadrados, que podem ser obtidas utilizando-se a função nls do R.

## 2.4 Predição dos efeitos aleatórios

A predição dos efeitos aleatórios no modelo (2.5) é realizada utilizando-se o método de Bayes empírico, que consiste em prever  $\mathbf{u}_i$  pela sua média condicional, dadas as respostas  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ . Pela suposição feita em (2.6), segue, por propriedades da distribuição normal multivariada (Muirhead 1982, cap.1), que

$$\mathbf{u}_i | \mathbf{y}_i \sim N_q(\mathbf{Q}\mathbf{Z}_i^\top \Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i)), \mathbf{Q} - \mathbf{Q}\mathbf{Z}_i^\top \Sigma_i^{-1} \mathbf{Z}_i \mathbf{Q}).$$

Para  $\Sigma_i$  fixada, a predição dos efeitos aleatórios  $\mathbf{u}_i$  usando-se o método de Bayes empírico é dada por

$$\hat{\mathbf{u}}_i = E[\widehat{\mathbf{u}}_i | \mathbf{y}_i] = \mathbf{Q}\mathbf{Z}_i^\top \Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i)). \quad (2.9)$$

Como  $\Sigma_i$  é desconhecida, é usual substituí-la por sua estimativa de máxima verossimilhança. Portanto, a predição de  $\hat{\mathbf{y}}_i$  é fornecida por

$$\begin{aligned}
\hat{\mathbf{y}}_i &= \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{Z}_i \hat{\mathbf{u}}_i \\
&= \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{Z}_i \hat{\mathbf{Q}} \mathbf{Z}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i)) \\
&= (\mathbf{I}_{m_i} - \mathbf{Z}_i \hat{\mathbf{Q}} \mathbf{Z}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1}) \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{Z}_i \hat{\mathbf{Q}} \mathbf{Z}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{y}_i \\
&= \hat{\sigma}^2 \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + (\mathbf{I}_{m_i} - \hat{\sigma}^2 \hat{\boldsymbol{\Sigma}}_i^{-1}) \mathbf{y}_i,
\end{aligned}$$

que pode ser interpretada como uma média ponderada do perfil populacional  $\boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i)$  e dos dados observados  $\mathbf{y}_i$ , com pesos  $\hat{\sigma}^2 \hat{\boldsymbol{\Sigma}}_i^{-1}$  e  $(\mathbf{I}_{m_i} - \hat{\sigma}^2 \hat{\boldsymbol{\Sigma}}_i^{-1})$ , respectivamente.

No capítulo seguinte é proposta uma generalização do modelo misto parcialmente não linear ao se adicionar uma função não paramétrica para modelar a média da variável resposta.

# Capítulo 3

## Modelo mistos semiparamétrico parcialmente não linear

Neste capítulo é proposta uma generalização do modelo desenvolvido por Vonesh & Carter (1992), a qual neste trabalho é designada modelo misto semiparamétrico parcialmente não linear. Na Seção 3.1 serão discutidas a motivação para esse modelo, possíveis aplicações e alguns trabalhos relacionados. Na Seção 3.2 o modelo é definido formalmente e aspectos relevantes são discutidos. Na Seção 3.3 é apresentada a representação matricial para o modelo, para, na seção 3.5, se construir o logaritmo da função de verossimilhança penalizada. Na Seção 3.6 é apresentado o processo de estimação dos parâmetros referentes aos efeitos fixos e da função não paramétrica.

### 3.1 Introdução

Modelos mistos não lineares, como apresentados no Capítulo 2, são ferramentas úteis para se estudarem dados correlacionados que apresentam relação não linear entre as covariáveis e a variável resposta. Porém, esses modelos podem ser muito restritivos e certos conjuntos de dados são muito complexos e informativos para serem totalmente modelados de forma paramétrica. Neste trabalho é proposto o modelo misto semiparamétrico parcialmente não linear (MMSPNL) (Machado & Russo 2013), que pode ser visto como uma generalização direta do modelo misto não linear apresentado por Vonesh & Carter (1992) e, portanto, uma generalização de vários outros modelos. Um MMSPNL assume que a média da variável resposta pode ser explicada por uma função paramétrica não linear, com relação aos parâmetros e às covariáveis, e uma função não paramétrica. Os parâmetros proporcionam um resumo interpretável dos dados, enquanto a função não paramétrica fornece flexibilidade para deixar que os dados “decidam” sobre componentes desconhecidos ou incertos, como a forma funcional da média da resposta ao longo do tempo. Os efeitos aleatórios, usados para modelar a estrutura de variâncias-covariâncias

dos dados, são incluídos de forma linear, o que proporciona a vantagem de se obter o modelo marginal de forma direta, sem que sejam necessárias integrações numéricas. A função não paramétrica é aproximada usando-se suavização por *B-splines*. A função de verossimilhança penalizada baseada no modelo marginal é utilizada para se estimarem todos os parâmetros e a função não paramétrica. A penalidade é baseada no operador de diferenças apresentado por Eilers & Marx (1996).

Há poucos trabalhos envolvendo modelos mistos semiparamétricos não lineares na literatura. Ke & Wang (2001) apresentam um modelo misto semiparamétrico não linear, em que a variável resposta é explicada por uma função que pode ser não linear com relação a parâmetros, função não paramétrica e efeitos aleatórios. Assim, torna-se necessário utilizar aproximações numéricas para se obter a verossimilhança para o modelo marginal. A complexidade desse modelo dificulta muito o processo de estimação e inferência. Liu & Wu (2008) apresentam uma aplicação desses modelos para dados de HIV com dados perdidos não ignoráveis e com erros de medição. Elmi et al. (2011) desenvolvem um método alternativo para ajustar o modelo apresentado por Ke & Wang (2001), baseando a suavização em termos de *B-splines*.

## 3.2 Especificação do modelo

Considere uma amostra consistindo de  $n$  sujeitos e  $m_i$  observações do  $i$ -ésimo sujeito ao longo do tempo. Seja  $y_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, m_i$ ) a  $j$ -ésima resposta do  $i$ -ésimo sujeito. O modelo misto semiparamétrico parcialmente não linear (MMSPNL) assume a seguinte forma

$$y_{ij} = \eta(\boldsymbol{\beta}, \mathbf{x}_{ij}) + f(t_{ij}) + \mathbf{Z}_{ij}^T \mathbf{u}_i + \varepsilon_{ij}, \quad (3.1)$$

em que  $\eta$  é uma função não linear de  $\boldsymbol{\beta}$  e  $\mathbf{x}_{ij}$ ,  $\mathbf{x}_{ij}$  é uma covariável que pode ser um escalar ou um vetor,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  é o vetor de parâmetros fixos e desconhecidos,  $\mathbf{Z}_{ij}$  é um vetor conhecido relacionado aos efeitos aleatórios,  $\mathbf{u}_i = (u_{i1}, \dots, u_{ir})^T$  é o vetor de coeficientes de efeitos aleatórios associado ao  $i$ -ésimo sujeito,  $\varepsilon_{ij}$  são os erros aleatórios,  $f$  é uma função real univariada do tempo e  $t_{ij}$  são os instantes de tempo.

No modelo (3.1) os efeitos fixos, constituídos pelas componentes paramétricas e não paramétricas, são utilizados para modelar o valor esperado da variável resposta  $y_{ij}$ . A função  $f$  deve ser considerada como um parâmetro do modelo (3.1) assim como  $\boldsymbol{\beta}$ . A diferença é que o último pertence a um espaço de dimensão finita (usualmente  $\mathbb{R}^p$ ) enquanto  $f$ , sendo uma função, pertence a um espaço de dimensão infinita. No processo de estimação é preciso especificar a qual espaço  $f$  pertence. A escolha desse espaço depende dos conhecimentos *a priori* sobre a função  $f$ , como o domínio, a forma funcional, restrições para garantir identificabilidade e propósitos da análise (Ke & Wang 2001). Uma possibilidade é considerar que  $f$  é uma função do espaço de Sobolev de

ordem 2, ou seja,  $f \in W^2[a, b]$ , em que

$$W^2[a, b] = \left\{ f : f \in C^1[a, b], \int_a^b \{f^{(2)}(x)\}^2 dx < \infty \right\}, \quad (3.2)$$

e  $C^1[a, b]$  representa o conjunto das funções definidas no intervalo  $[a, b]$  que possuem derivadas de primeira ordem contínuas.

Segundo Reinsch (1967) e Silverman (1985), as estimativas de máxima verossimilhança penalizada da função  $f$  pertencem a um espaço de dimensão finita. Especificamente, são polinômios cúbicos. Dessa forma, faz sentido considerar a expansão da função  $f$  em uma combinação linear de polinômios cúbicos.

Em particular, um método amplamente utilizado é considerar *B-splines* (ver Apêndice B) para realizar tal expansão (Eilers & Marx 1996, Elmi et al. 2011). Elas consistem em partes de polinômios, conectados de uma forma conveniente em valores especificados do domínio da função  $f$ , os nós. Uma vez dados os nós, as *B-splines* podem ser calculadas facilmente de maneira recursiva. Existem funções em linguagem R úteis para se construir, computacionalmente, as matrizes de representação dessas funções base. A escolha do número e posição dos nós é uma questão importante para se obter uma boa aproximação para a função  $f$ . Uma quantidade grande de nós leva a um sobreajuste dos dados, enquanto uma quantidade pequena deles leva a um subajuste dos mesmos. Para contornar esse problema, uma possibilidade é considerar um número relativamente grande de nós. Para prevenir sobreajuste, uma penalidade, com relação à forma funcional da curva ajustada, pode ser considerada (ver Seção 3.5). O uso de *B-splines* é justificado por serem elas uma boa aproximação da função  $f$ , estimada de forma não paramétrica, que apresentam eficiência computacional nas estimativas, não possuem efeitos de borda e permitem extensões diretas dos modelos de regressão paramétricos.

Neste trabalho, a expansão da função  $f$  é realizada com *B-splines* de grau 3 com  $l$  nós equidistantes,

$$f(t_{ij}) = \alpha_1 B_1(t_{ij}) + \dots + \alpha_l B_l(t_{ij}) \quad (3.3)$$

em que  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_l)^\top$  é um vetor desconhecido de coeficientes das *B-splines* para serem estimados.

Com relação aos efeitos aleatórios, esses são úteis para modelar as curvas específicas para cada indivíduo, assim como a estrutura de variâncias-covariâncias. Eles podem ser preditos utilizando-se o método de Bayes empírico.

A terminologia MMSPNL vem do fato de que o modelo (3.1) contém uma parte paramétrica representada pela função  $\eta$  que pode ser não linear com relação aos parâmetros  $\boldsymbol{\beta}$  e as covariáveis, uma parte não paramétrica representada pela função  $f$  e os efeitos aleatórios incorporados de forma linear ao modelo. O modelo (3.1) pode ser visto como uma generalização de vários modelos amplamente estudados nos últimos anos, como

pode ser observado por meio dos exemplos apresentados adiante. No entanto, pode ser considerado como uma generalização direta do modelo misto parcialmente não linear (2.5), desenvolvido por Vonesh & Carter (1992).

**Modelo misto parcialmente não linear.**

Quando  $f = 0$ , o modelo (3.1) assume a forma do modelo misto parcialmente não linear

$$y_{ij} = \eta(\boldsymbol{\beta}, \mathbf{x}_{ij}) + \mathbf{Z}_{ij}^\top \mathbf{u}_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (3.4)$$

apresentada no Capítulo 2. Esse modelo foi apresentado inicialmente por Vonesh & Carter (1992), considerando que os erros e efeitos aleatórios seguem uma distribuição normal. Russo (2010) generaliza esse trabalho supondo uma família de distribuições elípticas para as componentes aleatórias do modelo.

**Modelo misto linear**

Quando  $\eta(\boldsymbol{\beta}, \mathbf{x}_{ij}) = \mathbf{x}_{ij}^\top \boldsymbol{\beta}$  e  $f = 0$ , tem-se o amplamente estudado modelo misto linear proposto por Laird & Ware (1982),

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{Z}_{ij}^\top \mathbf{u}_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (3.5)$$

em que os erros e efeitos aleatórios são normalmente distribuídos.

**Modelo misto semiparamétrico linear**

Quando  $\eta(\mathbf{x}_{ij}, \boldsymbol{\beta}) = \mathbf{x}_{ij}^\top \boldsymbol{\beta}$  e  $\varepsilon_{ij} = z_i(t_{ij}) + \epsilon_{ij}$ , com  $z_i(t_{ij})$  sendo um processo estocástico independente de  $\epsilon_{ij}$ , o modelo (3.1) assume a seguinte forma

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + f(t_{ij}) + \mathbf{Z}_{ij}^\top \mathbf{u}_i + z_i(t_{ij}) + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (3.6)$$

que é o modelo estocástico misto semiparamétrico apresentado por Zhang et al. (1998), em que os erros e efeitos aleatórios são mutuamente independentes uns dos outros e seguem uma distribuição normal.

**Modelo misto não paramétrico**

Quando  $\boldsymbol{\beta} = 0$ , tem-se o modelo misto não paramétrico proposto por Wang (1998),

$$y_{ij} = \mathbf{Z}_{ij}^\top \mathbf{u}_i + f(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (3.7)$$

em que os erros e efeitos aleatórios são normalmente distribuídos.

### 3.3 Representação matricial

Em representação matricial, o modelo (3.1) para a variável resposta  $\mathbf{y}_i$  ( $m_i \times 1$ ) pode ser expresso como

$$\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{B}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad (3.8)$$

em que  $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^\top$ ,  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})^\top$ ,  $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{im_i})^\top$ ,  $\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) = (\eta(\boldsymbol{\beta}, \mathbf{x}_{i1}), \dots, \eta(\boldsymbol{\beta}, \mathbf{x}_{im_i}))^\top$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ ,  $\mathbf{u}_i = (u_{i1}, \dots, u_{ir})^\top$ ,  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  e os elementos da matriz de representação de *B-splines*  $\mathbf{B}_i$  ( $m_i \times l$ ) são  $b_{jl} = B_l(t_{ij})$  para o  $i$ -ésimo sujeito,  $i = 1, \dots, n$ .

Neste trabalho a função não paramétrica  $f$  é tratada de acordo com a metodologia desenvolvida por Eilers & Marx (1996), em que é realizada uma expansão em combinação linear de *B-splines* como apresentada na equação (3.3). Os autores fornecem funções, em linguagem S-Plus e MATLAB, úteis para se construir as matrizes  $\mathbf{B}_i$  ( $i = 1, \dots, n$ ). Green & Silverman (1994), no entanto, apresentam um método para estimar a função  $f$  de maneira não paramétrica, a qual é amplamente utilizada por muitos autores (Ibacache-Pulgar et al. 2012, Zhang et al. 1998).

Como dito anteriormente, a matriz  $\mathbf{Z}_i$  relacionada aos efeitos aleatórios é conhecida. Existem diversas estruturas que se podem considerar. Sejam  $\mathbf{x}_i$  as matrizes de covariáveis e  $\mathbf{1} = (1, \dots, 1)^\top$ , então  $\mathbf{Z}_i$  pode ser representada por

(i)  $\mathbf{Z}_i = \mathbf{1}$ ,

(ii)  $\mathbf{Z}_i = (\mathbf{1}, \mathbf{x}_i)$ ,

(iii)  $\mathbf{Z}_i = (\mathbf{1}, \mathbf{x}_i, \mathbf{x}_i^2)$  e

(iv)  $\mathbf{Z}_i = \left[ \frac{\partial \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i)}{\partial \boldsymbol{\beta}} \right] \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$ ,

em que  $\tilde{\boldsymbol{\beta}}$  representam as estimativas de mínimos quadrados de  $\boldsymbol{\beta}$  para o modelo sem efeitos aleatórios,  $i = 1, \dots, n$ .

### 3.4 Pressupostos do modelo

Neste trabalho é assumido que  $\mathbf{u}_i$  e  $\boldsymbol{\varepsilon}_i$  são mutuamente independentes e seguem uma distribuição normal multivariada. Assim, a distribuição conjunta do vetor aleatório  $(\mathbf{y}_i^\top, \mathbf{u}_i^\top, \boldsymbol{\varepsilon}_i^\top)^\top$  é fornecida por

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{u}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \stackrel{ind.}{\sim} N_{2m_i+r} \left( \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{B}_i \boldsymbol{\alpha} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_i & \mathbf{Z}_i \mathbf{Q} & \mathbf{V}_i \\ \mathbf{Q} \mathbf{Z}_i^\top & \mathbf{Q} & \mathbf{0} \\ \mathbf{V}_i & \mathbf{0} & \mathbf{V}_i \end{pmatrix} \right), \quad (3.9)$$

$i = 1, \dots, n$ , em que as matrizes  $\Sigma_i = \mathbf{Z}_i \mathbf{Q} \mathbf{Z}_i^\top + \mathbf{V}_i$ ,  $\mathbf{Q}$ ,  $\mathbf{V}_i$  e  $\mathbf{Z}_i \mathbf{Q}$  são matrizes de variâncias-covariâncias  $\text{Var}(\mathbf{y}_i)$ ,  $\text{Var}(\mathbf{u}_i)$ ,  $\text{Var}(\boldsymbol{\varepsilon}_i)$  e  $\text{Cov}(\mathbf{y}_i, \mathbf{u}_i)$ , respectivamente. Nesse caso é fácil trabalhar com o modelo marginal para se preservar a média da variável resposta sem que sejam necessárias integrações numéricas, sendo essa uma grande vantagem desse modelo. Ademais, modela o mecanismo gerador dos dados por meio de uma função não linear e permite maior flexibilidade para explicar a média da variável resposta pela componente não paramétrica. Além disso, ao modelar a estrutura de correlação dos dados por meio dos efeitos aleatórios, adicionados de forma linear ao modelo, tem-se que o modelo marginal é obtido de maneira exata utilizando propriedades da distribuição normal multivariada,

$$\mathbf{y}_i \stackrel{ind.}{\sim} N_{m_i}(\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{B}_i \boldsymbol{\alpha}, \Sigma_i),$$

$i = 1, \dots, n$ .

### 3.5 Função de log-verossimilhança penalizada

Considere que as matrizes de variâncias-covariâncias  $\Sigma_i = \mathbf{Z}_i \mathbf{Q} \mathbf{Z}_i^\top + \mathbf{V}_i$ , são positivas definidas, em que  $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\tau})$  e  $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\delta})$  e dependem de um número finito de parâmetros, tais que  $\boldsymbol{\tau} \in \mathbb{R}^{d_\tau}$  e  $\boldsymbol{\delta} \in \mathbb{R}^{d_\delta}$ ,  $i = 1, \dots, n$ . Assim, o vetor de parâmetros a ser estimado no modelo misto semiparamétrico parcialmente não linear será  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\tau}^\top, \boldsymbol{\delta}^\top)^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$ , cujo espaço paramétrico associado é dado pelo conjunto

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^{p^*} \mid \boldsymbol{\beta} \in \Theta_\beta, \boldsymbol{\alpha} \in \Theta_\alpha, \boldsymbol{\gamma} \in \Theta_\gamma\},$$

em que  $p^* = p + l + d_\gamma$ ,  $\Theta_\beta \subseteq \mathbb{R}^p$ ,  $\Theta_\alpha \subseteq \mathbb{R}^l$  e  $\Theta_\gamma \subseteq \mathbb{R}^{d_\gamma}$  denotam os espaços paramétricos associados, respectivamente, aos efeitos fixos paramétricos, aos efeitos fixos não paramétricos e às componentes de variâncias-covariâncias. Assim, o logaritmo da função de verossimilhança para  $\boldsymbol{\theta}$  é, a menos de uma constante, dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}), \quad (3.10)$$

em que

$$L_i = -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) - \mathbf{B}_i \boldsymbol{\alpha}]^\top \Sigma_i^{-1} [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) - \mathbf{B}_i \boldsymbol{\alpha}]$$

é o logaritmo da verossimilhança associada ao  $i$ -ésimo indivíduo.

A maximização direta de (3.10) sem impor restrições à função  $f$  pode causar sobreajuste aos dados e inidentificabilidade de  $\boldsymbol{\beta}$  (Green 1987, Ibacache-Pulgar et al. 2012). Para tornar o ajuste menos flexível, um procedimento é incorporar em (3.10) uma função de penalidade,  $J(f)$ , construída para considerar características da função  $f$  como,



por exemplo, a suavidade. Assim, obtém-se a função de log-verossimilhança penalizada,

$$L_p(\boldsymbol{\theta}, \lambda) = L(\boldsymbol{\theta}) + \lambda J(f), \quad (3.11)$$

em que  $\lambda > 0$  é denominado parâmetro de suavização, que controla o peso relativo da função de penalidade na estimação de  $f$ .

Uma importante questão para se obter um bom ajuste é a escolha do parâmetro de suavização. No processo de estimação é comum propor uma rede de valores para  $\lambda$  e escolher seu valor ótimo baseado em algum critério de informação (Eilers & Marx 1996). Outra alternativa é considerar  $\lambda$  como um parâmetro adicional à função de verossimilhança penalizada. Zhang et al. (1998) utilizam esse método para estimar o parâmetro de suavização no modelo misto semiparamétrico linear. No entanto, realizar esse procedimento no caso não linear pode ser um problema complexo. Na Seção 3.6 o processo de estimação de  $\lambda$  será abordado de maneira mais aprofundada.

Com relação à função de penalidade, existem diversas alternativas (Goodd & Gaskins 1971, Shen 1997). Uma amplamente utilizada é a penalidade baseada na segunda derivada da função  $f$  (Green & Silverman 1994, Ibacache-Pulgar et al. 2012, O'sullivan et al. 1986),

$$J(f) = \int_a^b [f^{(2)}(x)]^2 dx, \quad (3.12)$$

para o caso em que  $f$  pertence ao espaço de Sobolev de ordem 2,

$$W^2[a, b] = \left\{ f : f \in C^1[a, b], \int_a^b \{f^{(2)}(x)\}^2 dx < \infty \right\}. \quad (3.13)$$

Segundo Green & Silverman (1994), é possível expressar a equação (3.12) como

$$J(f) = \int_a^b [f^{(2)}(x)]^2 dx = \mathbf{f}^\top \mathbf{M} \mathbf{f} \quad (3.14)$$

em que  $\mathbf{f}$  é um vetor de parâmetros associados à função  $f$  e  $\mathbf{M}_{s \times s}$  é uma matriz com coeficientes reais, definida positiva, que depende somente dos nós (Green & Silverman 1994). Observe que essa função de penalidade influencia diretamente na suavidade da função  $f$ . A grande vantagem de se trabalhar com esse caso é que a estimativa de máxima verossimilhança penalizada de  $f$  pertence a um espaço de dimensão finita, sendo especificamente, um polinômio cúbico.

Como apontado anteriormente neste trabalho, a função de suavização é uma combinação linear de *B-splines* cúbicas, fato que reduz o problema de estimar  $f$  em um espaço de dimensão infinita para um de dimensão finita. Seguimos Eilers & Marx (1996) e consideramos a função de penalidade baseada em diferenças finitas dos coeficientes de *B-splines* adjacentes, que fornece a função de log-verossimilhança penalizada para o

$i$ -ésimo indivíduo,

$$L_{p_i}(\boldsymbol{\theta}, \lambda) = L_i(\boldsymbol{\theta}) - \frac{\lambda}{2n} \boldsymbol{\alpha}^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\alpha}, \quad (3.15)$$

para  $i = 1, \dots, n$ , em que  $\mathbf{D}_k$  é a matriz de representação do operador diferença (de coeficientes de *B-splines* adjacentes) e  $k$  é a ordem das diferenças. Assim, a função de log-verossimilhança,  $L_p(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \lambda)$ , é fornecida por

$$L_p(\boldsymbol{\theta}, \lambda) = L(\boldsymbol{\theta}) - \frac{\lambda}{2} \boldsymbol{\alpha}^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\alpha}. \quad (3.16)$$

Nesse caso, em que a suavização é realizada por meio de *B-splines* e a função de penalidade baseada no operador de diferenças  $\mathbf{D}_k$ , temos a conhecida suavização com *P-splines*. Eilers & Marx (1996) justificam utilizar a penalidade baseada no operador de diferenças por ser uma boa aproximação discreta da integral do quadrado da  $k$ -ésima derivada da função de suavização. Computacionalmente, essa matriz pode ser calculada utilizando-se a função `diff()` no R.

### 3.6 Processo de estimação

Considere, para  $\lambda$  fixo, a função escore penalizada dada por

$$\mathbf{U}_p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta}}, \quad (3.17)$$

em que  $\partial L_p(\boldsymbol{\theta}, \lambda)/\partial \boldsymbol{\theta}$  representa a primeira derivada parcial da função de log-verossimilhança penalizada com relação ao vetor  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\tau}^\top, \boldsymbol{\delta}^\top)^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$ . Os componentes da função escore penalizada são fornecidas pelas equações

$$\begin{aligned} \mathbf{U}_p^\beta &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i, \\ \mathbf{U}_p^\alpha &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^n \mathbf{B}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i - \lambda \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\alpha} \quad \text{e} \\ \mathbf{U}_p^\gamma &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\gamma}} = (U_{\gamma_1}, \dots, U_{\gamma_{d_\gamma}})^\top, \text{ com} \\ U_{\gamma_j} &= -\frac{1}{2} \sum_{i=1}^n \left\{ \text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \right] - \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i^{-1}(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \right\}, \end{aligned}$$

em que  $\mathbf{J}_i = \partial \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}^\top$ ,  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $\dot{\boldsymbol{\Sigma}}_i(j) = \partial \boldsymbol{\Sigma}_i / \partial \gamma_j$ , para  $j = 1, \dots, d_\gamma$ ,  $i = 1, \dots, n$ , e  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{d_\gamma})^\top = (\boldsymbol{\tau}^\top, \boldsymbol{\delta}^\top)^\top$ .

A matriz de informação de Fisher penalizada para o vetor de parâmetros  $\boldsymbol{\theta}$  é definida por

$$\mathbf{K}_p(\boldsymbol{\theta}) = -\mathbb{E} \left\{ \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\}, \quad (3.18)$$

em que  $\partial^2 L_p(\boldsymbol{\theta}, \lambda) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$  representa as derivadas parciais da função de log-verossimilhança

penalizada e é fornecida pela matriz

$$\mathbf{K}_p^{\theta\theta} = \begin{bmatrix} \mathbf{K}_p^{\beta\beta} & \mathbf{K}_p^{\beta\alpha} & \mathbf{0} \\ \mathbf{K}_p^{\beta\alpha\top} & \mathbf{K}_p^{\alpha\alpha} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_p^{\gamma\gamma} \end{bmatrix},$$

em que

$$\begin{aligned} \mathbf{K}_p^{\beta\beta} &= \sum_{i=1}^n \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{J}_i, \\ \mathbf{K}_p^{\alpha\alpha} &= \sum_{i=1}^n \mathbf{B}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_i + \lambda \mathbf{D}_k^\top \mathbf{D}_k, \\ \mathbf{K}_p^{\beta\alpha} &= \sum_{i=1}^n \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_i \text{ e} \\ \mathbf{K}_p^{\gamma\gamma} &= \sum_{i=1}^n K_{i\gamma}, \text{ cujo } (q, s)\text{-ésimo elemento é dado por} \\ K_{i\gamma,qs} &= \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) \right]. \end{aligned}$$

Um algoritmo iterativo para se obterem as estimativas de máxima verossimilhança penalizada para  $\boldsymbol{\theta}$  usando o método score de Fisher é dado por

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}_{(m+1)} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}_{(m)} + \begin{bmatrix} \mathbf{K}_p^{\beta\beta} & \mathbf{K}_p^{\beta\alpha} \\ \mathbf{K}_p^{\beta\alpha\top} & \mathbf{K}_p^{\alpha\alpha} \end{bmatrix}_{(m)}^{-1} \begin{pmatrix} \mathbf{U}_p^\beta \\ \mathbf{U}_p^\alpha \end{pmatrix}_{(m)}$$

$$\hat{\boldsymbol{\gamma}}_{(m+1)} = \hat{\boldsymbol{\gamma}}_{(m)} + (\mathbf{K}_p^{\gamma\gamma})_{(m)}^{-1} \mathbf{U}_{p(m)}^\gamma, \quad m = 0, 1, 2, \dots$$

com  $\mathbf{K}_p^{\beta\beta}$ ,  $\mathbf{K}_p^{\alpha\alpha}$ ,  $\mathbf{K}_p^{\beta\alpha}$ ,  $\mathbf{K}_p^{\gamma\gamma}$ ,  $\mathbf{U}_p^\beta$ ,  $\mathbf{U}_p^\alpha$  e  $\mathbf{U}_p^\gamma$  como apresentado anteriormente. Os valores iniciais para o algoritmo podem ser, por exemplo, as estimativas de mínimos quadrados.

### 3.6.1 O parâmetro de suavização

O parâmetro de suavização  $\lambda$  controla o peso relativo da função de penalidade, ou equivalentemente, controla a suavidade da curva estimada. Segundo Eilers & Marx (1996), uma questão importante em qualquer técnica de suavização é a escolha da quantidade ótima de suavização. O processo de estimação de  $\lambda$  consiste em propor uma rede grande de valores para esse parâmetro e escolher seu valor ótimo segundo algum critério de informação. Neste trabalho será considerado o critério de informação Akaike (AIC). A definição de AIC para o MMSPNL é dada por

$$AIC(\lambda) = 2k - 2L_p(\hat{\boldsymbol{\theta}}, \lambda), \quad (3.19)$$

em que  $k$  é o número de parâmetros no modelo e  $L_p(\hat{\boldsymbol{\theta}}, \lambda)$  é o valor da função de log-verossimilhança penalizada (3.5), avaliada em sua estimativa de máxima verossimilhança para cada  $\lambda$  fixo.

Tal procedimento é equivalente a comparar o maior valor da função de verossimilhança, uma vez que  $k$  é fixo para todos os valores considerados para  $\lambda$ .

### 3.6.2 Predição dos efeitos aleatórios e da variável resposta

A predição dos efeitos aleatórios no modelo (3.1) é realizada utilizando-se o método de Bayes empírico, que consiste em prever  $\mathbf{u}_i$  pela sua média condicional, dadas as respostas  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ . Devido aos pressupostos de distribuição para as componentes aleatórias, segue por propriedades da distribuição normal multivariada (Muirhead 1982, cap. 1), que

$$\mathbf{u}_i | \mathbf{y}_i \sim N_q \left( \mathbf{QZ}_i^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha}), \mathbf{Q} - \mathbf{QZ}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{Q} \right).$$

Para  $\boldsymbol{\Sigma}_i$  fixada, a predição dos efeitos aleatórios  $\mathbf{u}_i$  é dada por

$$\begin{aligned} \hat{\mathbf{u}}_i &= E[\widehat{\mathbf{u}}_i | \mathbf{y}_i] \\ &= \mathbf{QZ}_i^\top \boldsymbol{\Sigma}_i^{-1} \left( \mathbf{y}_i - \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) - \mathbf{B}_i \hat{\boldsymbol{\alpha}} \right). \end{aligned} \quad (3.20)$$

Como  $\boldsymbol{\Sigma}_i$  é desconhecida, é usual substituí-la por sua estimativa de máxima verossimilhança penalizada. Portanto, a predição de  $\hat{\mathbf{y}}_i$  é fornecida por

$$\begin{aligned} \hat{\mathbf{y}}_i &= \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{B}_i \hat{\boldsymbol{\alpha}} + \mathbf{Z}_i \hat{\mathbf{u}}_i \\ &= \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1} \left( \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{B}_i \hat{\boldsymbol{\alpha}} \right) + (\mathbf{I}_{m_i} - \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}) \mathbf{y}_i. \end{aligned}$$

que pode ser interpretado como uma média ponderada do perfil populacional  $\boldsymbol{\eta}(\hat{\boldsymbol{\beta}}, \mathbf{x}_i) + \mathbf{B}_i \hat{\boldsymbol{\alpha}}$  e dos dados observados  $\mathbf{y}_i$ , com pesos  $\widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}$  e  $(\mathbf{I}_{m_i} - \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1})$ , respectivamente.

### 3.6.3 Erros padrão das estimativas

Uma importante questão é como determinar a matriz de variâncias-covariâncias das estimativas de máxima verossimilhança penalizada. De acordo com Ibacache-Pulgar et al. (2012), pode-se aproximar essa matriz pela inversa da matriz de Fisher penalizada. Assim, a matriz de variâncias-covariâncias assintótica aproximada de  $\hat{\boldsymbol{\theta}}$  é fornecida por

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) \approx [\mathbf{K}_p^{\theta\theta}(\hat{\boldsymbol{\theta}})]^{-1}.$$

Finalizando a discussão sobre o MMSPNL, no próximo capítulo é apresentado o método de influência local para esse modelo.

# Capítulo 4

## Influência local

Neste capítulo é desenvolvido o método de influência local para o modelo misto semiparamétrico parcialmente não linear, baseado na curvatura normal conformal. Na Seção 4.1 são discutidos alguns trabalhos relacionados ao assunto. Na Seção 4.2 é apresentado o método de influência local. Na Seção 4.3 são apresentadas as deduções da curvatura normal de Cook, da curvatura normal conformal e da matriz de informação. Finalmente, na Seção 4.4 é apresentado o método de influência local baseado em três esquemas de perturbações: ponderação de casos, perturbação na matriz escala e perturbação da variável resposta.

### 4.1 Introdução

O método de influência local consiste em verificar as variações que pequenas perturbações nos dados ou nos pressupostos do modelo podem causar nas estimativas dos parâmetros. Se essas variações não forem representativas, o desconhecimento do modelo preciso não causará problemas para extrair informações de um conjunto de dados. Cook (1986) propôs um enfoque geométrico para a verificação das suposições do modelo, assim como a identificação de dados aberrantes e/ou influentes, ao estudar o efeito de pequenas perturbações no modelo (ou dados) usando uma medida de influência adequada (Osório et al. 2007). Lee & Xu (2004) desenvolveram análises de influência local para modelos mistos não lineares assumindo normalidade. É possível citar alguns exemplos de estudos de influência local supondo distribuições elípticas para os componentes aleatórios do modelo. Osório et al. (2007) desenvolveram o método para modelos lineares aplicados a dados longitudinais, Russo et al. (2009) desenvolveram o método para modelos mistos parcialmente não lineares e Ibacache-Pulgar et al. (2012) desenvolveram o método para modelos lineares mistos semiparamétricos. Na seção seguinte é desenvolvido o método de influência local para o modelo misto semiparamétrico parcialmente não linear.

## 4.2 Método de influência local

Para desenvolver o método de influência local baseado na função de verossimilhança penalizada para o MMSPNL, considere o parâmetro de suavização  $\lambda$  fixo. Para um conjunto de dados específico, seja  $L_p(\boldsymbol{\theta}, \lambda)$  a função de log-verossimilhança penalizada correspondente ao modelo postulado, em que  $\boldsymbol{\theta}$  é um vetor  $p \times 1$  de parâmetros desconhecidos em um espaço paramétrico  $\Theta$ , as perturbações no modelo são introduzidas pelo vetor  $\boldsymbol{\omega}$  de dimensão  $q \times 1$  que pertence a um subconjunto aberto  $\Omega$  de  $\mathbb{R}^q$ . Seja  $L_p(\boldsymbol{\theta}, \lambda|\boldsymbol{\omega})$  a função de log-verossimilhança penalizada correspondente ao modelo perturbado para um dado  $\boldsymbol{\omega}$  pertencente a  $\Omega$ . Assume-se que existe  $\boldsymbol{\omega}_0$  em  $\Omega$  tal que  $L_p(\boldsymbol{\theta}, \lambda) = L_p(\boldsymbol{\theta}, \lambda|\boldsymbol{\omega}_0)$ ,  $\boldsymbol{\theta} \in \Theta$ . Sejam  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  os estimadores de máxima verossimilhança penalizada obtidos como máximos de  $L_p(\boldsymbol{\theta}, \lambda)$  e  $L_p(\boldsymbol{\theta}, \lambda|\boldsymbol{\omega})$ , respectivamente, e assumamos que  $L_p(\boldsymbol{\theta}, \lambda|\boldsymbol{\omega})$  é duas vezes continuamente diferenciável com relação ao vetor  $(\boldsymbol{\theta}, \boldsymbol{\omega})^\top$ . Para verificar a influência das perturbações nas estimativas de máxima verossimilhança penalizada  $\hat{\boldsymbol{\theta}}$ , considere o afastamento da verossimilhança penalizada

$$LD(\boldsymbol{\omega}) = 2[L_p(\hat{\boldsymbol{\theta}}, \lambda) - L_p(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}, \lambda)]. \quad (4.1)$$

O gráfico de  $LD(\boldsymbol{\omega})$  contra  $\boldsymbol{\omega}$  contém informações essenciais sobre a influência dos esquemas de perturbações. Considere o gráfico formado pelos valores do vetor  $(q+1) \times 1$

$$\boldsymbol{\varphi}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, LD(\boldsymbol{\omega}))^\top \quad (4.2)$$

denominado gráfico de influência. Cook (1986) propôs usar a curvatura normal para caracterizar o comportamento de um gráfico de influência em torno de  $\boldsymbol{\omega}_0$ .

Observe-se que um gráfico de influência é uma superfície em  $\mathbb{R}^{q+1}$  e a noção de curvatura pode ser entendida como segue. Deseja-se descrever como a superfície  $\boldsymbol{\varphi}(\boldsymbol{\omega})$  se diferencia de seu plano tangente em  $\boldsymbol{\omega}_0$ . Essa descrição pode ser obtida estudando-se a curvatura de certas curvas da superfície passando por  $\boldsymbol{\varphi}(\boldsymbol{\omega}_0)$ . Visualizando em  $\mathbb{R}^3$ , essas curvas são seções normais formadas pelas interseções da superfície com planos contendo o vetor que é normal ao plano tangente em  $\boldsymbol{\omega}_0$ . As curvaturas dessas seções normais são chamadas de curvaturas normais (do Carmo 1976) e são essas curvaturas que Cook (1986) propôs para estudar o comportamento de um gráfico de influência em torno de  $\boldsymbol{\omega}_0$ .

Como  $LD(\boldsymbol{\omega})$  atinge um mínimo local em  $\boldsymbol{\omega}_0$ , para construir uma seção normal, considere uma reta em  $\Omega$  passando por  $\boldsymbol{\omega}_0$ . Essa reta pode ser representada por

$$\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\boldsymbol{l}, \quad (4.3)$$

em que  $a \in \mathbb{R}$ ,  $\boldsymbol{l}$  é um vetor fixado pertencente a  $\mathbb{R}^q$  e  $\|\boldsymbol{l}\| = 1$ . Essa linha gera a linha

projetada no gráfico de influência  $\varphi(\boldsymbol{\omega})$  passando por  $\varphi(\boldsymbol{\omega}_0)$ . Cada direção  $\boldsymbol{l}$  especifica uma linha projetada e cada linha projetada corresponde a uma seção normal. Cook (1986) mostrou que a curvatura normal na direção unitária  $\boldsymbol{l}$  toma a forma

$$C_{\boldsymbol{l}} = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}}_p)^{-1} \boldsymbol{\Delta}_p \boldsymbol{l}|, \quad (4.4)$$

em que  $\|\boldsymbol{l}\| = 1$ ,  $\ddot{\mathbf{L}}_p$  é a matriz de informação observada penalizada para o modelo postulado ( $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ ), cujos elementos são dados pela relação

$$\ddot{\mathbf{L}}_p = \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (4.5)$$

enquanto  $\boldsymbol{\Delta}_p$  é uma matrix  $p \times q$  com elementos

$$\boldsymbol{\Delta}_{p_{ij}} = \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \theta_i \partial \omega_j} \quad (4.6)$$

avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ ,  $i = 1, \dots, p$  e  $j = 1, \dots, q$ . Para estudar o comportamento de  $\varphi(\boldsymbol{\omega})$ , Cook (1986) sugeriu considerar a direção  $\boldsymbol{l}_{max}$  que corresponde à máxima curvatura  $C_{\boldsymbol{l}_{max}}(\boldsymbol{\theta})$ . O gráfico de  $\boldsymbol{l}_{max}$  contra a ordem das observações pode revelar as observações que têm uma influência desproporcional sob pequenas perturbações em  $LD(\boldsymbol{\omega})$ . Observe que  $\boldsymbol{l}_{max}$  é o autovetor referente ao maior autovalor absoluto da matriz  $\boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}}_p)^{-1} \boldsymbol{\Delta}_p$ .

Considere a partição de  $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$  e suponha que se deseja avaliar a influência somente sobre o subvetor  $\boldsymbol{\theta}_1$ . Nesse caso, a curvatura normal na direção de  $\boldsymbol{l}$  é dada por  $C_{\boldsymbol{l}}(\boldsymbol{\theta}) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}} - \mathbf{B}_1) \boldsymbol{\Delta}_p \boldsymbol{l}|$ , em que

$$\mathbf{B}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22} \end{pmatrix},$$

sendo que  $\ddot{\mathbf{L}}_{22}$  representa a matriz de informação observada relacionada ao subvetor  $\boldsymbol{\theta}_2$ . O gráfico de índices do autovetor correspondente ao maior autovalor de  $\boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}} - \mathbf{B}_1) \boldsymbol{\Delta}_p$  contra a ordem das observações pode indicar os pontos com grande influência em  $\hat{\boldsymbol{\theta}}_1$ .

As dificuldades em trabalhar com a curvatura normal surgem na medida em que a curvatura normal pode tomar qualquer valor e não é invariante sob mudanças uniformes de escala. Como resultado, não existe um critério objetivo para se julgar a magnitude das curvaturas normais e o tamanho relativo das componentes de direções correspondentes a grandes curvaturas normais. Para contornar esse problema, Poon & Poon (1999)

propuseram utilizar a curvatura normal conformal na direção unitária  $\mathbf{l}$

$$B_{\mathbf{l}}(\boldsymbol{\theta}) = \frac{-\mathbf{l}^\top \boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}}_p)^{-1} \boldsymbol{\Delta}_p \mathbf{l}}{\sqrt{\text{tr}(\boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}}_p)^{-1} \boldsymbol{\Delta}_p)^2}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad (4.7)$$

que é uma transformação injetora da curvatura normal  $C_{\mathbf{l}}(\boldsymbol{\theta})$ .

Seja  $\Omega$  o conjunto de perturbações. Uma reparametrização é uma aplicação suave  $\phi : \Omega \rightarrow \Omega'$  do domínio  $\Omega$  para um novo domínio  $\Omega'$  com a mesma dimensão, tal que a matriz Jacobiana de  $\phi$  é não singular em  $\Omega$ . Em particular, essa é uma aplicação injetora e, portanto, é invertível (Poon & Poon 1999). Nesse trabalho, reparametrizações são consideradas modificações dos esquemas de perturbações. Uma matriz quadrada  $\mathbf{M}$  é uma matriz conformal se existe um número real positivo  $\xi$  tal que  $\mathbf{M}\mathbf{M}^\top = \xi\mathbf{I}_n$ , em que  $\mathbf{I}_n$  é a matriz identidade de ordem  $n$ . Uma reparametrização é conformal em  $\boldsymbol{\omega}_0$  se sua matriz Jacobiana é uma matriz conformal nesse ponto.

A curvatura normal conformal possui propriedades importantes que serão brevemente discutidas a seguir. Se uma reparametrização de  $\Omega$  é conformal em um ponto crítico  $\boldsymbol{\omega}_0$  do gráfico de  $LD$  sobre  $\Omega$ , então a curvatura normal conformal em qualquer direção em  $\boldsymbol{\omega}_0$  é invariante sob reparametrizações. Além disso, para qualquer direção  $\mathbf{l}$ ,  $B_{\mathbf{l}}$  satisfaz a condição que  $0 \leq |B_{\mathbf{l}}| \leq 1$ . Esse resultado implica que  $B_{\mathbf{l}}$  é uma medida padronizada e, portanto, interpretar sua magnitude se torna uma tarefa mais simples do que interpretar a magnitude da curvatura  $C_{\mathbf{l}}$ .

Seja  $\{\mathbf{e}_1, \dots, \mathbf{e}_s\}$  um conjunto de autovetores ortonormais da matriz

$$\boldsymbol{\Delta}_p^\top (\ddot{\mathbf{L}}_p)^{-1} \boldsymbol{\Delta}_p,$$

com autovalores padronizados  $\nu_1, \dots, \nu_s$ . Então  $B_{\mathbf{e}_i}$  é igual ao autovalor padronizado  $\nu_i$ ,  $i = 1, \dots, s$ . Como resultado, tem-se que  $\sum_i B_{\mathbf{e}_i}^2 = 1$ .

Para se avaliar a influência local de um esquema de perturbação, Poon & Poon (1999) definem que um autovetor é  $q$  influente se  $|B_{\mathbf{e}_i}| \geq q/\sqrt{n}$ . Seja  $\mathbf{E}_i$  o vetor coluna em  $\mathbb{R}^n$  cuja  $i$ -ésima entrada é igual a 1 e todas as outras iguais a 0,  $i = 1, \dots, n$ , denominados vetores de perturbação básica do espaço de perturbações. Para se analisar a influência dos vetores de perturbação básica em todos os autovetores, considere o agrupamento dos valores absolutos dos autovalores normalizados em ordem decrescente

$$\nu_{max} = \nu_1 \geq \dots \geq \nu_k \geq q/\sqrt{n} > \nu_{k+1} \geq \dots \geq \nu_s \geq 0$$

e denote por  $a_{ij}$  o  $j$ -ésimo elemento do autovetor normalizado correspondente a  $\nu_i$ ,  $i = 1, \dots, s$ . A contribuição agregada da  $j$ -ésima perturbação básica para todos  $q$



autovetores influentes é fornecida por

$$m(q)_j = \sqrt{\sum_{i=1}^k \nu_i a_{ij}^2}.$$

Quando  $q = 0$  é possível avaliar a contribuição agregada de todos os autovetores. Ademais, existe  $q$  suficientemente grande que possibilita avaliar apenas a influência na direção do autovetor correspondente ao maior autovalor, ou seja, a influência na direção de  $\mathbf{l}_{max}$ .

Para julgar a magnitude de  $m(q)$  uma possibilidade é considerar  $\overline{m}(q) + c^* sm(q)$ , em que  $\overline{m}(q)$  e  $sm(q)$  representam a média e o desvio padrão de  $m(q)$ , respectivamente, e  $c^*$  é uma constante selecionada (Lee & Xu 2004, Russo 2010).

### 4.3 Derivação da curvatura

Nesta seção são apresentadas a matriz de informação observada penalizada  $-\ddot{\mathbf{L}}_p(\boldsymbol{\theta})$  e a matriz de perturbação penalizada  $\boldsymbol{\Delta}_p(\boldsymbol{\theta})$  para diferentes esquemas de perturbação. Consideram-se os esquemas de ponderação de casos, perturbação da matriz escala e perturbação na variável resposta. Para tanto, considere o parâmetro de suavização,  $\lambda$ , fixo e denote por  $\dot{\mathbf{Q}}(j) = \partial \mathbf{Q} / \partial \tau_j$ ,  $\ddot{\mathbf{Q}}(j, k) = \partial^2 \mathbf{Q} / \partial \tau_j \partial \tau_k$ ,  $\dot{\mathbf{V}}_i(t) = \partial \mathbf{V}_i / \partial \delta_t$  e  $\ddot{\mathbf{V}}(t, k) = \partial^2 \mathbf{V}_i / \partial \delta_t \partial \delta_k$ , as derivadas parciais das matrizes de escala  $\mathbf{Q}$  e  $\mathbf{V}_i$ .

#### 4.3.1 Matriz de informação

A matriz de informação observada penalizada é dada por

$$-\ddot{\mathbf{L}}_p(\boldsymbol{\theta}) = -\sum_{i=1}^n \ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}),$$

em que

$$\ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}) = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\mathbf{L}}_{p_i}^{\beta\beta} & \ddot{\mathbf{L}}_{p_i}^{\beta\alpha} & \ddot{\mathbf{L}}_{p_i}^{\beta\tau} & \ddot{\mathbf{L}}_{p_i}^{\beta\delta} \\ \ddot{\mathbf{L}}_{p_i}^{\alpha\beta} & \ddot{\mathbf{L}}_{p_i}^{\alpha\alpha} & \ddot{\mathbf{L}}_{p_i}^{\alpha\tau} & \ddot{\mathbf{L}}_{p_i}^{\alpha\delta} \\ \ddot{\mathbf{L}}_{p_i}^{\tau\beta} & \ddot{\mathbf{L}}_{p_i}^{\tau\alpha} & \ddot{\mathbf{L}}_{p_i}^{\tau\tau} & \ddot{\mathbf{L}}_{p_i}^{\tau\delta} \\ \ddot{\mathbf{L}}_{p_i}^{\delta\beta} & \ddot{\mathbf{L}}_{p_i}^{\delta\alpha} & \ddot{\mathbf{L}}_{p_i}^{\delta\tau} & \ddot{\mathbf{L}}_{p_i}^{\delta\delta} \end{pmatrix}$$

e para cada  $i = 1, \dots, n$ ,

$$\ddot{\mathbf{L}}_{p_i}^{\beta\beta} = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{J}_i,$$

$$\ddot{\mathbf{L}}_{p_i}^{\beta\alpha} = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}^\top} = \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_i,$$

$\ddot{\mathbf{L}}_{p_i}^{\beta\tau}$  é uma matriz cuja  $j$ -ésima coluna,  $j = 1, \dots, d_\tau$ , é fornecida por

$$\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \tau_j} = -\mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,$$

$\ddot{\mathbf{L}}_{p_i}^{\beta\delta}$  é uma matriz cuja  $t$ -ésima coluna,  $t = 1, \dots, d_\delta$ , é fornecida por

$$\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \delta_t} = -\mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(t) \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,$$

$$\ddot{\mathbf{L}}_{p_i}^{\alpha\alpha} = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} = -[\mathbf{B}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_i + \lambda \mathbf{D}_k^\top \mathbf{D}_k],$$

$\ddot{\mathbf{L}}_{p_i}^{\alpha\tau}$  é uma matriz cuja  $j$ -ésima coluna,  $j = 1, \dots, d_\tau$ , é fornecida por

$$\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\alpha} \partial \tau_j} = -\mathbf{B}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,$$

$\ddot{\mathbf{L}}_{p_i}^{\alpha\delta}$  é uma matriz cuja  $t$ -ésima coluna,  $t = 1, \dots, d_\delta$ , é fornecida por

$$\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\alpha} \partial \delta_t} = -\mathbf{B}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(t) \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,$$

$\ddot{\mathbf{L}}_{p_i}^{\tau\tau}$  é uma matriz com elementos fornecidos por

$$\begin{aligned} \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \tau_j \partial \tau_k} &= \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i [\dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(k) - \ddot{\mathbf{Q}}(j, k)] \mathbf{U}_i^\top] \\ &+ \frac{1}{2} \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i [\ddot{\mathbf{Q}}(j, k) - \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(k) - \dot{\mathbf{Q}}(k) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j)] \\ &\times \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \end{aligned}$$

$\ddot{\mathbf{L}}_{p_i}^{\tau\delta}$  é uma matriz com elementos fornecidos por

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \tau_j \partial \delta_t} &= \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(t) \mathbf{U}_i^\top] \\ &\quad - \frac{1}{2} \mathbf{r}_i \boldsymbol{\Sigma}_i^{-1} [\dot{\mathbf{V}}(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(t) \mathbf{U}_i^\top + \mathbf{U}_i \dot{\mathbf{Q}}(t) \mathbf{U}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(j)] \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,\end{aligned}$$

$\ddot{\mathbf{L}}_{p_i}^{\delta\delta}$  é uma matriz com elementos fornecidos por

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \lambda)}{\partial \delta_t \partial \delta_k} &= \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}_i^{-1} [\dot{\mathbf{V}}(t) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(k) - \ddot{\mathbf{V}}(t, k)]] \\ &\quad + \frac{1}{2} \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} [\ddot{\mathbf{V}}(t, k) + \dot{\mathbf{V}}(t) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(t) - \dot{\mathbf{V}}(k) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}(t)] \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,\end{aligned}$$

em que  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $j = 1, \dots, d_\tau$  e  $t = 1, \dots, d_\delta$ .

## 4.4 Matriz de perturbação

A matriz  $\Delta_p$  de perturbação de casos penalizada é obtida pela seguinte relação

$$\Delta_p = \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \Delta_p^\beta(\boldsymbol{\theta}) \\ \Delta_p^\alpha(\boldsymbol{\theta}) \\ \Delta_p^\tau(\boldsymbol{\theta}) \\ \Delta_p^\delta(\boldsymbol{\theta}) \end{pmatrix}$$

em que  $\Delta_p^\beta(\boldsymbol{\theta}) = \partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top$ ,  $\Delta_p^\alpha(\boldsymbol{\theta}) = \partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\omega}^\top$ ,  $\Delta_p^\tau(\boldsymbol{\theta}) = \partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) / \partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top$  e  $\Delta_p^\delta(\boldsymbol{\theta}) = \partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) / \partial \boldsymbol{\delta} \partial \boldsymbol{\omega}^\top$ , sendo os elementos de cada matriz avaliados em suas estimativas de máxima verossimilhança penalizada  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e no vetor de não perturbação  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ .

A seguir apresentaremos três esquemas de perturbação, ponderação de casos, perturbação na matriz escala e perturbação na variável resposta.

### 4.4.1 Ponderação de casos

O esquema de perturbação de ponderação de casos consiste em perturbar multiplicativamente o logaritmo da função de verossimilhança referente a cada grupo de observações. O objetivo é identificar possíveis grupos de medidas influentes em cada caso. Sejam  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$  o vetor de pesos, com  $0 \leq \omega_i \leq 1$ ,  $i = 1, \dots, n$  e o vetor de não perturbação  $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ . O logaritmo da função de verossimilhança penalizada

com os pesos atribuídos é fornecida por

$$L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\lambda}{2} \boldsymbol{\alpha}^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\alpha}, \quad (4.8)$$

em que  $L_i = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) - \mathbf{B}_i \boldsymbol{\alpha}]^\top \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}) - \mathbf{B}_i \boldsymbol{\alpha}]$ . Diferenciando  $L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})$  com relação a  $\boldsymbol{\theta}$  e  $\omega_i$ , obtêm-se as equações

$$\begin{aligned} \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{J}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i, \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\alpha} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{B}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \tau_j \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \text{tr}[\hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top] + \frac{1}{2} \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i \text{ e} \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \delta_t \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \text{tr}[\hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(t)] + \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(t) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i, \end{aligned}$$

em que  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d_\tau$  e  $t = 1, \dots, d_\delta$ .

#### 4.4.2 Perturbação na matriz escala

O esquema de perturbação na matriz escala permite avaliar a influência que exercem as observações na estrutura de escala e no vetor das componentes de variância. Essa perturbação pode ser realizada tomando

$$\mathbf{y}_i \sim N_{m_i}(\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{B}_i \boldsymbol{\alpha}, \omega_i^{-1} \boldsymbol{\Sigma})$$

em que  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top \in \mathbb{R}^n - \{\mathbf{0}\}$ , o vetor de não perturbação é dado por  $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$  e  $i = 1, \dots, n$ . Diferenciando com relação a  $\boldsymbol{\theta}$  e  $\boldsymbol{\omega}$ , obtêm-se as equações

$$\begin{aligned} \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{J}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i, \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\alpha} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{B}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i, \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \tau_j \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \frac{1}{2} \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{U}}_i \dot{\mathbf{Q}}(j) \hat{\mathbf{U}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i \text{ e} \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \delta_t \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \frac{1}{2} \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(t) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{r}}_i, \end{aligned}$$

em que  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $j = 1, \dots, d_\tau$ ,  $t = 1, \dots, d_\delta$  e  $i = 1, \dots, n$ .

### 4.4.3 Perturbação na variável resposta

Um esquema de perturbação na variável resposta consiste em considerar o vetor  $\mathbf{y}_{i\omega} = \mathbf{y}_i + \boldsymbol{\omega}_i$ , em que  $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{im_i})^\top$  é o vetor de perturbação para o  $i$ -ésimo grupo e  $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$  é o vetor de não perturbação. O objetivo é avaliar a sensibilidade das estimativas quando são introduzidas pequenas perturbações nas componentes de cada vetor de respostas. O logaritmo da função de verossimilhança penalizada do modelo perturbado é fornecido por

$$L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta} | \boldsymbol{\omega}_i) - \frac{\lambda}{2} \boldsymbol{\alpha}^\top \mathbf{D}_k^\top \mathbf{D}_k \boldsymbol{\alpha}, \quad (4.9)$$

em que  $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}_i) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} (\mathbf{y}_{i\omega} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_{i\omega} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $i = 1, \dots, n$ . Diferenciando  $L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})$  com relação a  $\boldsymbol{\theta}$  e  $\boldsymbol{\omega}_i$ , obtem-se

$$\begin{aligned} \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{J}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1}, \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\omega}_i^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{B}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1}, \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \tau_j \partial \boldsymbol{\omega}_i^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{U}_i \dot{\mathbf{Q}}(j) \mathbf{U}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \text{ e} \\ \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \delta_t \partial \boldsymbol{\omega}_i^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{r}}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(t) \hat{\boldsymbol{\Sigma}}_i^{-1}, \end{aligned}$$

em que  $\mathbf{r}_i = (\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{B}_i \boldsymbol{\alpha})$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d_\tau$  e  $t = 1, \dots, d_\delta$ .

No próximo capítulo será apresentada a aplicação do MMSPNL para os dados de concentração de theophylline, assim como uma análise de diagnóstico baseada na identificação de observações aberrantes utilizando o método de influência local, desenvolvido neste capítulo, e uma análise de resíduos.

# Capítulo 5

## Aplicação

Neste capítulo é apresentada a aplicação do modelo misto semiparamétrico parcialmente não linear aos dados de concentração de *theophylline*.

### 5.1 Dados de concentração de *theophylline*

Em um experimento descrito em Davidian & Giltinan (1995), o antiasmático *theophylline* foi administrado via oral a 12 sujeitos e concentrações dessa substância foram medidas em 11 instantes de tempo (ver Apêndice A). O objetivo de se utilizar a metodologia desenvolvida neste trabalho com este conjunto de dados é identificar possíveis ganhos na modelagem estatística, ao se considerar o modelo semiparamétrico (equação 3.1), ao invés do modelo paramétrico (equação 2.5). A Figura 5.1 representa o diagrama de dispersão dos dados.

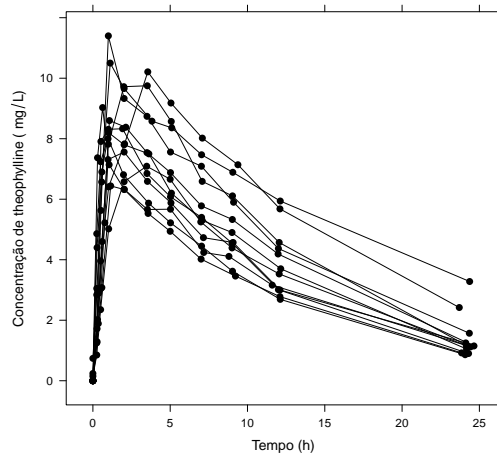


Figura 5.1: Diagrama de dispersão dos dados de concentração de *theophylline*.

As covariáveis são os instantes de tempo  $t_{ij}$  (h) e as doses  $d_i$  (mg/kg) administradas ao  $i$ -ésimo indivíduo. A variável resposta é a concentração de *theophylline*  $conc_{ij}$  (mg/L),

$i = 1, \dots, 12$  e  $j = 1, \dots, 11$ . Note-se que existe uma correlação intrínseca entre as observações do mesmo indivíduo, representadas pelas linhas no gráfico. Além disso, os dados apresentam relação não linear entre a concentração de *theophylline* e o tempo. A Figura 5.2 representa o *boxplot* dos dados.

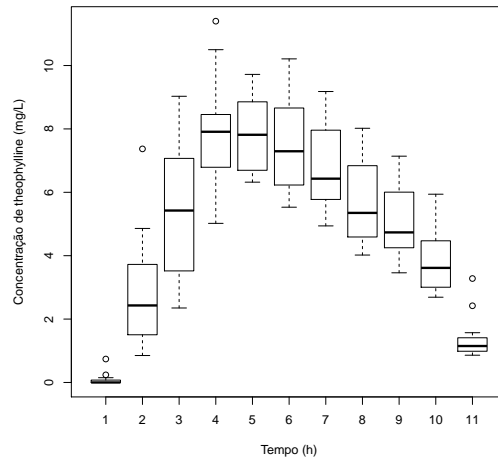


Figura 5.2: Boxplot dos dados de concentração de *theophylline*.

É possível observar pelo gráfico da Figura 5.2 que, inicialmente, os dados apresentam pouca variação entre as concentrações da substância nos indivíduos. Essa variação aumenta nas medidas seguintes e novamente fica pequena na última medida. Os valores extremos referentes à primeira medida são os indivíduos 1 e 10, que receberam doses iniciais de 4,02 e 5,50 (mg/kg), respectivamente. As doses iniciais variam de 3,10 (mg/kg) a 5,86 (mg/kg). Já o valor extremo referente à segunda medida é o indivíduo 9 que recebeu a menor dose 3,10 (mg/kg). Isso pode indicar que esse indivíduo possui alta taxa de absorção da substância. O valor extremo referente à quarta medida é o indivíduo 5 que recebeu a maior dose 5,86 (mg/kg). Finalmente, os valores extremos referentes à última medida são os indivíduos 1 e 10.

## 5.2 Modelo proposto

Usualmente nessa aplicação se considera a seguinte função não linear para modelar o mecanismo gerador dos dados (Pinheiro & Bates 2002)

$$\eta(\boldsymbol{\beta}|t_{ij}, d_i) = d_i \exp(lK_e + lK_a - lC_l) \frac{(\exp(-e^{lK_e} t_{ij}) - \exp(-e^{lK_a} t_{ij}))}{e^{lK_a} - e^{lK_e}}, \quad (5.1)$$

em que  $i = 1, \dots, 12$ ,  $j = 1, \dots, 11$  e  $\boldsymbol{\beta} = (lK_a, lK_e, lC_l)^\top$  é o vetor de parâmetros fixos. Tem-se que  $lK_a$ ,  $lK_e$  e  $lC_l$  representam o logaritmo da absorção (1/h), o logaritmo da eliminação (1/h) e o logaritmo da depuração plasmática (L/kg), respectivamente. As

quantidades  $d_i$  (mg/kg) representam a dose administrada ao  $i$ -ésimo indivíduo. Denotando por  $y_{ij}$  a concentração de *theophylline* do indivíduo  $i$  no tempo  $j$  ( $i = 1, \dots, 12$ ;  $j = 1, \dots, 11$ ), o modelo proposto é definido por

$$y_{ij} = \eta(\boldsymbol{\beta}|t_{ij}, d_i) + f(t_{ij}) + \mathbf{Z}_{ij}\mathbf{u}_i + \varepsilon_{ij} \quad (5.2)$$

em que  $\eta(\boldsymbol{\beta}|t_{ij}, d_i)$  é como apresentado em (5.1),  $f$  é uma função arbitrária que depende do tempo,  $\mathbf{Z}_{ij}$  é a  $j$ -ésima linha da matriz (5.3),  $\mathbf{u}_i$  é o vetor de efeitos aleatórios e  $\varepsilon_{ij}$  são os erros aleatórios.

Neste trabalho, a matriz associada aos efeitos aleatórios é fornecida por

$$\mathbf{Z}_i = \left[ \frac{\partial \eta(\boldsymbol{\beta}, \mathbf{t}_i)}{\partial lK_e}, \frac{\partial \eta(\boldsymbol{\beta}, \mathbf{t}_i)}{\partial lK_a}, \frac{\partial \eta(\boldsymbol{\beta}, \mathbf{t}_i)}{\partial lC_l} \right] \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}, \quad (5.3)$$

com  $\tilde{\boldsymbol{\beta}}$  sendo as estimativas de mínimos quadrados de  $\boldsymbol{\beta} = (lK_a, lK_e, lC_l)^\top$ , para  $i = 1, \dots, 12$ .

Munido da notação adotada neste trabalho, verifica-se que as matrizes  $\mathbf{Z}_i$  possuem dimensões  $11 \times 3$ , pois existem 11 observações no tempo para cada indivíduo. Isso implica que a matriz  $\mathbf{Q}$  de variâncias-covariâncias dos efeitos aleatórios possui dimensão  $3 \times 3$ . Para um modelo mais parcimonioso, considere que  $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\tau})$  seja uma matriz diagonal com elementos do vetor  $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$ , o que significa que os efeitos aleatórios são descorrelacionados. Além disso, considere que os erros aleatórios são homoscedásticos e possuem matriz de variâncias-covariâncias  $\mathbf{V}_i = \sigma^2 \mathbf{I}_{11}$ , para  $i = 1, \dots, 12$ . Para modelar a função não paramétrica foi utilizado um total de 10 nós, já que, para cada indivíduo, há 11 observações. Assim, foi necessário estimar 10 coeficientes de *B-splines*. Note-se que no processo de estimação, esses coeficientes são tratados como parâmetros de efeitos fixos. Entretanto, eles não possuem interpretação como os parâmetros associados à função paramétrica, que modela a relação funcional entre a covariável tempo e a variável resposta concentração de *theophylline*. Assim, o vetor de parâmetros a ser estimado tem 17 componentes, fornecido por  $\boldsymbol{\theta} = (lK_a, lK_e, lC_l, \alpha_1, \dots, \alpha_{10}, \sigma^2, \tau_1, \tau_2, \tau_3)$ , cujo espaço paramétrico associado é dado por  $\boldsymbol{\Theta} = \{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbb{R}^3 \times \mathbb{R}^{10} \times \mathbb{R}^+ \times \mathbb{R}^3\}$ . O número total de parâmetros a ser estimado é menor do que o número de observações, já que são utilizados dados de todos os indivíduos para estimar os parâmetros que modelam a média da variável resposta. Resumindo, temos 132 observações para estimar 17 parâmetros.

O parâmetro de suavização foi escolhido usando o critério AIC. Como o número de parâmetros é fixo, o AIC se reduz a escolher o  $\lambda$  que maximiza a log-verossimilhança penalizada. O melhor modelo, segundo o critério AIC, dentre os ajustados foi obtido com  $\hat{\lambda} = 0,01$ , com função de log-verossimilhança penalizada igual a -181.53. Além disso, foi ajustado o modelo paramétrico para desenvolver comparações entre esse e o modelo semiparamétrico. Os resultados são apresentados na Figura 5.3 e na Tabela



5.1. A Figura 5.4 representa o gráfico de índice dos efeitos aleatórios preditos para o modelo semiparamétrico. Os resultados para o caso paramétrico foram semelhantes e serão omitidos.

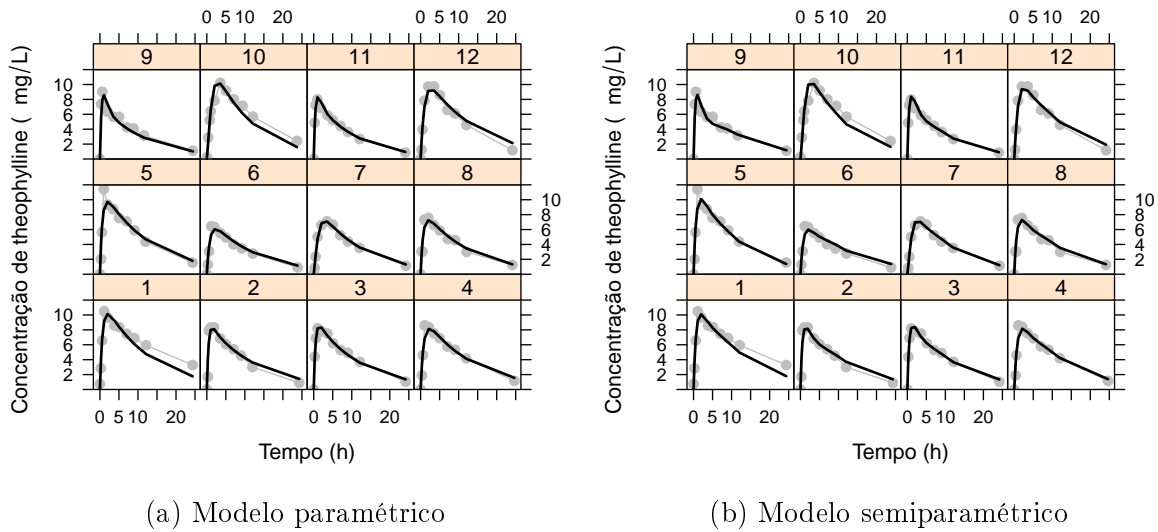


Figura 5.3: Modelos paramétrico e semiparamétrico ajustados

Tabela 5.1: Estimadores e seus respectivos erros padrão para os modelos paramétrico e semiparamétrico.

Parâmetros	Paramétrico		Semiparamétrico ( $\lambda = 0.01$ )	
	Estimativas	Erros padrão	Estimativas	Erros padrão
$lK_e$	-2,5209	0,0488	-1,9439	0,3703
$lK_a$	0,4514	0,1868	0,41180	0,2286
$lC_l$	-3,2628	0,0613	-2,7521	0,3992
$\sigma^2$	0,5895	0,0815	0,56213	0,0781
$\tau_1$	-0,0086	0,0075	-0,0069	0,0075
$\tau_2$	0,4098	0,1764	0,47219	0,1987
$\tau_3$	0,0286	0,0119	0,02617	0,0112

Pela Figura 5.3, o ajuste obtido para o caso semiparamétrico visualmente difere pouco do ajuste paramétrico. Em ambos os casos o modelo se ajusta bem aos dados da maior parte dos indivíduos. Porém, nos indivíduos 1 e 10, os modelos tendem a subestimar as curvas individuais, ou seja, os modelos prevêm um decaimento mais rápido da substância no organismo do que o decaimento realmente observado. Já nos indivíduos 2 e 6, os modelos tendem a superestimar as curvas individuais, ou seja, os modelos prevêm um decaimento mais lento da substância no organismo do que o decaimento realmente observado. No indivíduo 5, os modelos apresentam um ajuste coerente em relação aos dados, com exceção de um ponto extremo que destoa das observações de

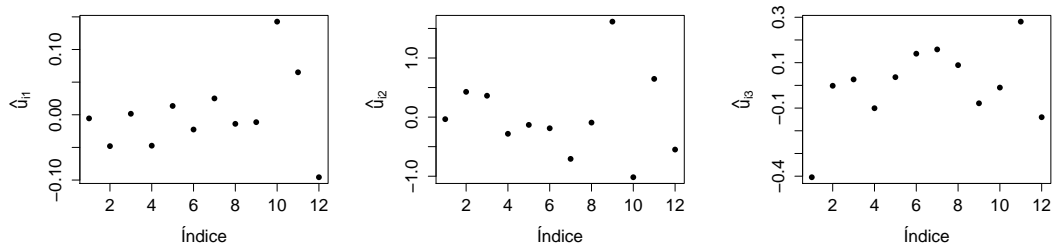


Figura 5.4: Gráfico de índice dos efeitos aleatórios preditos para o modelo semiparamétrico

todos os outros indivíduos. Além disso, a Tabela 5.1 mostra que os erros padrão no modelo paramétrico são, na maior parte dos casos, menores que os obtidos para o modelo semiparamétrico. A interpretação dos parâmetros da função não linear é semelhante, já que as estimativas dos parâmetros possuem sinais iguais.

A Figura 5.4 ilustra a variabilidade dos efeitos aleatórios preditos para o modelo semiparamétrico, que permitem o ajuste das curvas específicas para cada indivíduo. Os gráficos de efeitos aleatórios preditos para o modelo paramétrico são semelhantes ao modelo semiparamétrico e serão omitidos.

### 5.3 Análise de resíduos

Para identificar observações aberrantes com relação aos modelos paramétricos e semiparamétricos, uma possibilidade é desenvolver análise de resíduos considerando a distância de Mahalanobis,

$$\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad (5.4)$$

em que  $\boldsymbol{\mu}_i = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}_i) + \mathbf{B}_i \boldsymbol{\alpha}$ ,  $i = 1, \dots, n$ . Para o caso normal,  $\delta_i$  segue uma distribuição qui-quadrado com  $m_i$  graus de liberdade. Assim, é possível considerar percentis de  $\chi_{m_i}^2(\xi)$ , em que  $0 < \xi < 1$ , para construir pontos de corte para identificação de pontos aberrantes (Salgado 2006). A Figura 5.5 representa a distância de Mahalanobis para os modelos ajustados. O ponto de corte foi construído considerando-se o percentil  $\xi = 0,975$ .

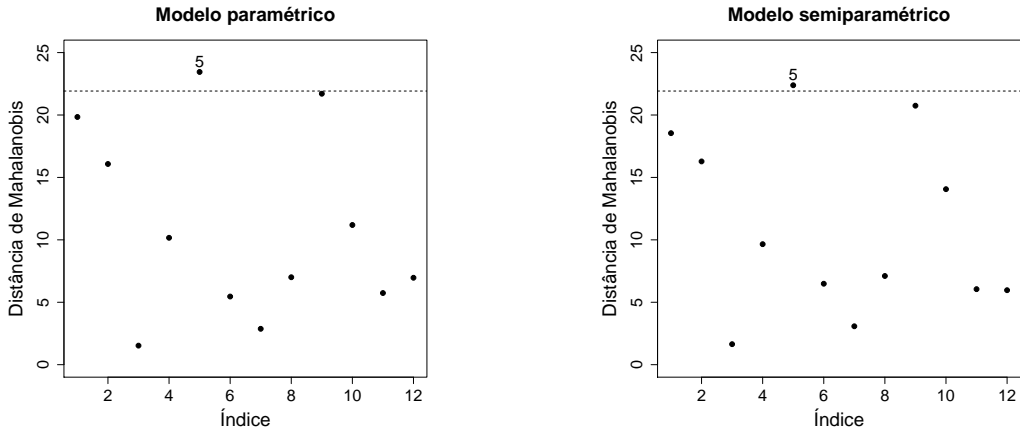


Figura 5.5: Distância de Mahalanobis para os modelos paramétrico e semiparamétrico

Os gráficos indicam o indivíduo 5 como possível observação aberrante. Ele possui a quarta maior medida da substância e recebeu a maior dose. As curvas ajustadas com os modelos paramétricos e semiparamétricos subestimam essa medida.

Para verificar a adequabilidade dos pressupostos dos modelos desenvolvidos, uma possibilidade seria considerar gráficos dos resíduos condicionais, definidos por

$$rc_{ij} = y_{ij} - \eta(\hat{\beta}, x_{ij}) - \hat{f}(t_{ij}) - \mathbf{Z}_{ij}\hat{\mathbf{u}}_i. \quad (5.5)$$

Os gráficos da Figura 5.6 representam os resíduos condicionais padronizados para os modelos paramétrico e semiparamétrico. Eles mostram que, para ambos os modelos ajustados, todas as observações ficam dentro do intervalo estabelecido  $(-1.96, 1.96)$ . Além disso, o comportamento aleatório em torno da origem indica um ajuste razoavelmente bom do modelo.

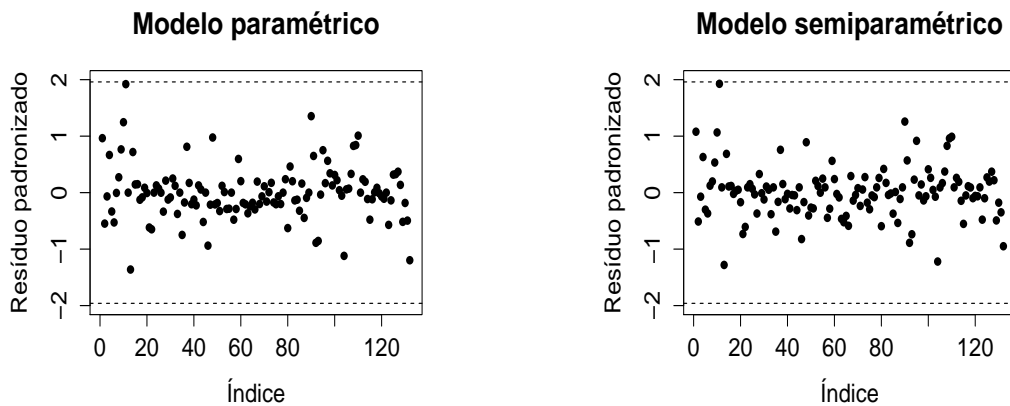


Figura 5.6: Resíduos condicionais padronizados para os modelos paramétrico e semiparamétrico

A Figura 5.7 representa os gráficos de quantis para os modelos paramétrico e semiparamétrico. Eles indicam que, em ambos os casos, os resíduos padronizados se aproximam dos quantis teóricos da distribuição normal, já que os pontos estão distribuídos próximos da reta identidade, o que é um indicio de um ajuste razoável do modelo.

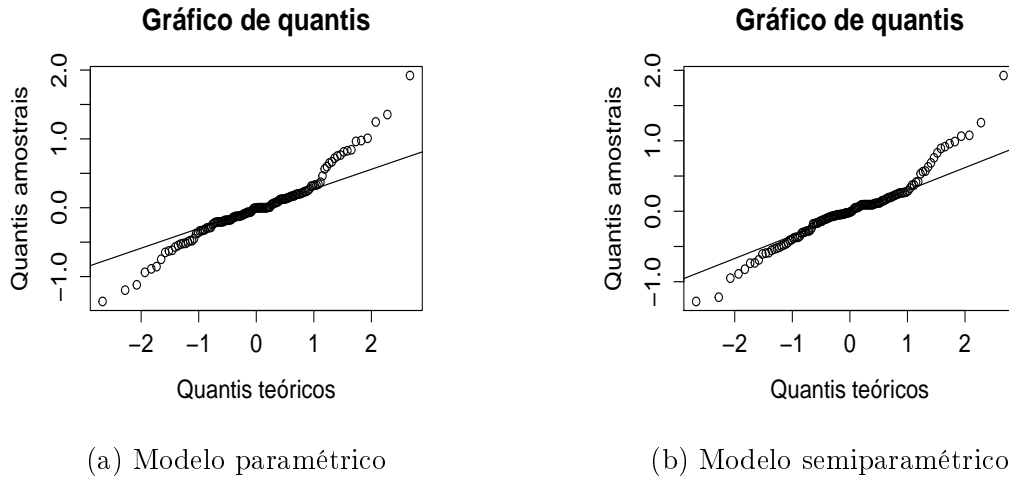


Figura 5.7: Gráficos de quantis dos resíduos condicionais padronizados.

## 5.4 Diagnóstico de Influência

Para realizar o estudo de influência local para os dados de concentração de *theophylline*, considere inicialmente o esquema ponderação de casos. Considerando  $q = 0$  é possível avaliar a influência agregada de todas as direções de influência. Considerando  $q = 2$  é possível avaliar a influência na direção de maior curvatura  $l_{max}$  (Figura 5.8). A medida  $\bar{m}(q) + 2sm(q)$  será utilizada para definir o ponto de corte para  $m(q)$ .

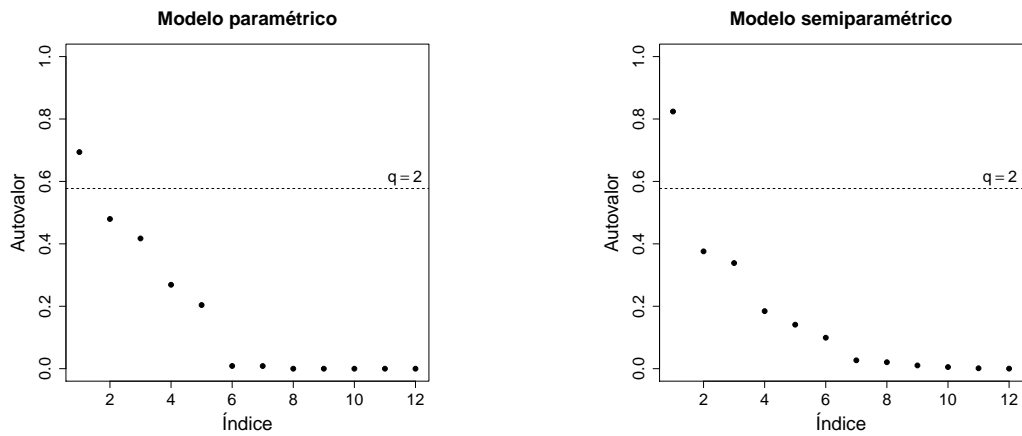


Figura 5.8: Autovalores relativos às direções de influência sob o esquema de ponderação de casos

Ao avaliar a contribuição agregada de todas as direções de influência em  $\hat{\theta}$ , por meio do gráfico de índices de  $m(0)$ , nenhuma observação se destacou como influente em ambos os modelos paramétrico e semiparamétrico (Figura 5.9).

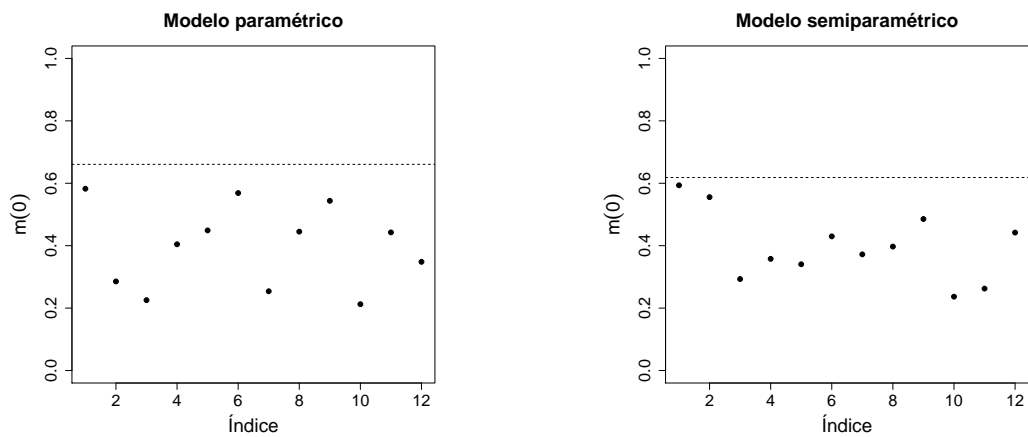


Figura 5.9: Gráficos de índices de  $m(0)$  para  $\hat{\theta}$  sob o esquema de ponderação de casos.

Ao avaliar a influência na direção de  $\mathbf{l}_{max}$  de maior curvatura, a observação 1 se destacou como provável ponto influente no modelo semiparamétrico e nenhum ponto se destacou no modelo paramétrico (Figura 5.10).

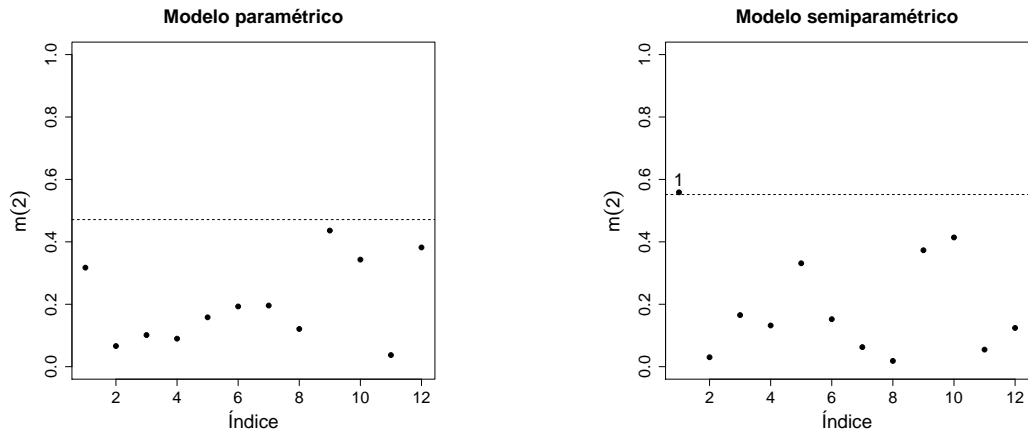


Figura 5.10: Gráficos de índices de  $m(2)$  para  $\hat{\theta}$  sob o esquema de ponderação de casos.

Com relação ao esquema perturbação na matriz escala, considerando  $q = 0$ , é possível avaliar a influência agregada de todas as direções de influência e, quando  $q = 2$  é possível avaliar a influência na direção de maior curvatura  $l_{max}$  (Figura 5.11). A medida  $\bar{m}(q) + 2sm(q)$  será utilizada para definir o ponto de corte para  $m(q)$ .

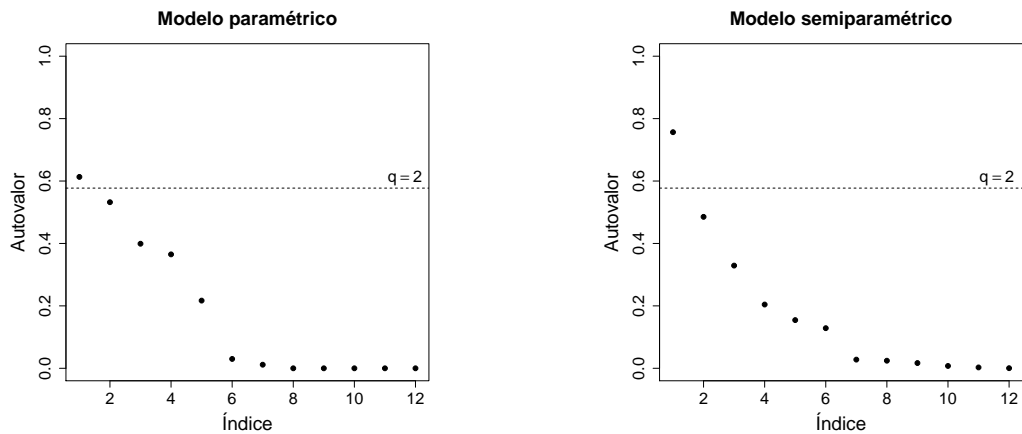


Figura 5.11: Autovalores relativos às direções de influência sob o esquema de perturbação na matriz escala.

Ao avaliar a contribuição agregada de todas as direções de influência em  $\hat{\theta}$ , por meio do gráfico de índices de  $m(0)$  sob o esquema perturbação na matriz escala, nenhuma observação se destacou como influente sob os modelos paramétrico e semiparamétrico (Figura 5.12).

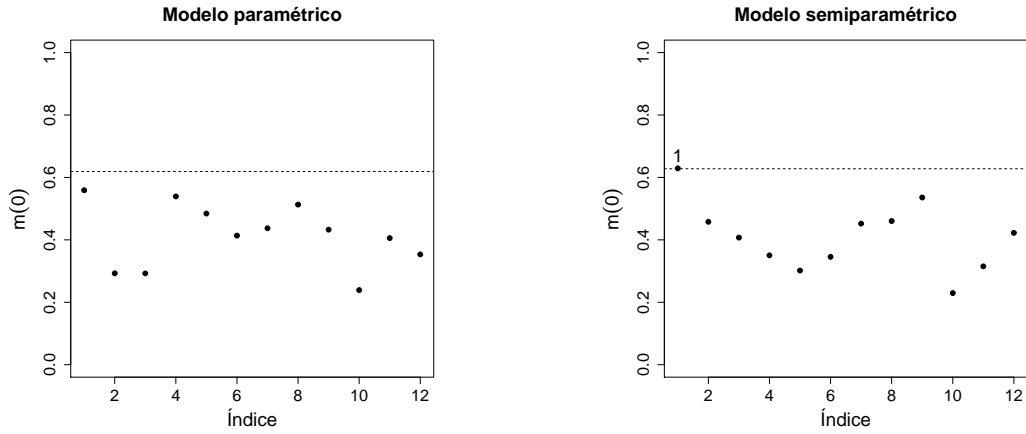


Figura 5.12: Gráficos de índices de  $m(0)$  para  $\hat{\theta}$  sob o esquema perturbação na matriz escala.

Ao avaliar a influência na direção de  $\mathbf{l}_{max}$  de maior curvatura sob o esquema perturbação na matriz escala, a observação 1 se destacou como provável ponto influente no modelo semiparamétrico e nenhum ponto se destacou no modelo paramétrico (Figura 5.13).

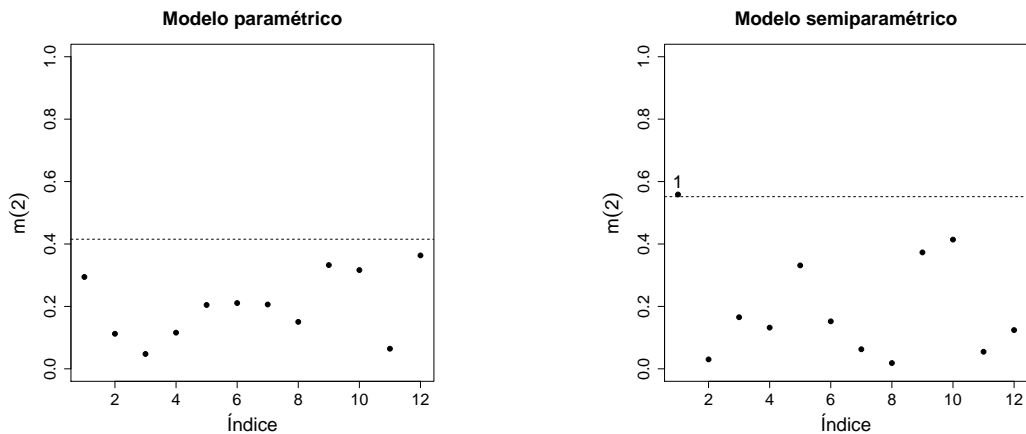


Figura 5.13: Gráficos de índices de  $m(2)$  para  $\hat{\theta}$  sob o esquema de perturbação na matriz escala.

Com relação ao esquema perturbação na variável resposta, considerando  $q = 0$  é possível avaliar a influência agregada de todas as direções de influência e, quando  $q = 6$  é possível avaliar a influência na direção de maior curvatura  $\mathbf{l}_{max}$  (Figura 5.14). A medida  $\bar{m}(q) + 3sm(q)$  será utilizada para definir o ponto de corte para  $m(q)$ .

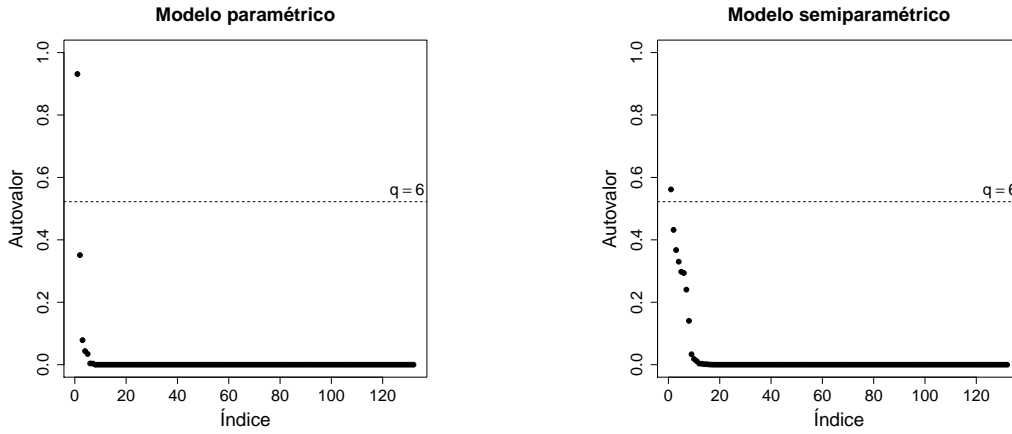


Figura 5.14: Autovalores relativos às direções de influência sob o esquema de perturbação na variável resposta

Ao avaliar a contribuição agregada de todas as direções de influência em  $\hat{\theta}$ , por meio do gráfico de índices de  $m(0)$  sob o esquema perturbação na variável resposta, a medida 4 se destacou como possível ponto influente sob o modelo paramétrico e as medidas 8, 12 e 124 se destacaram como influentes sob o modelo semiparamétrico (Figura 5.15). As medidas 4 e 8 são referentes ao indivíduo 1, a medida 12 referente ao indivíduo 2 e a medida 124 referente ao indivíduo 12.

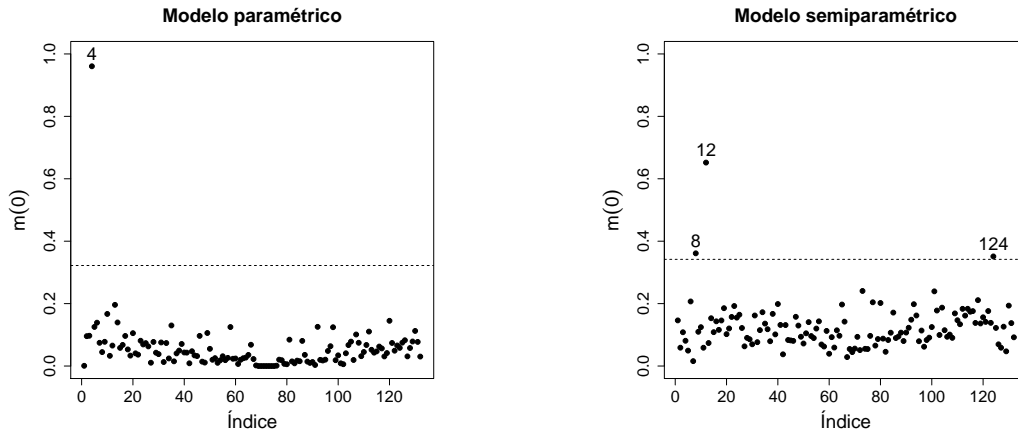


Figura 5.15: Gráficos de índices de  $m(0)$  para  $\hat{\theta}$  sob o esquema de perturbação na variável resposta.

Ao avaliar a influência na direção de  $l_{max}$  de maior curvatura sob o esquema perturbação na variável resposta, a medida 55, referente ao indivíduo 5, se destacou como provável ponto influente no modelo paramétrico e nenhum ponto se destacou no modelo semiparamétrico (Figura 5.16).



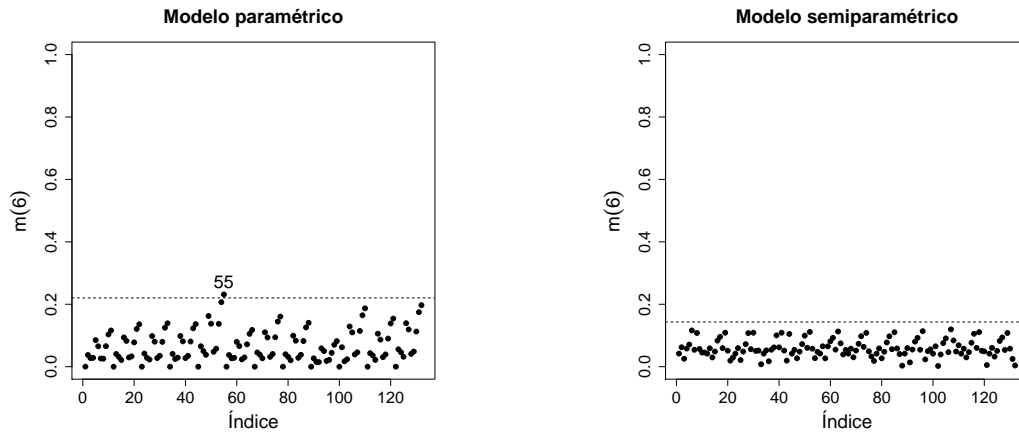


Figura 5.16: Gráficos de índices de  $m(6)$  para  $\hat{\theta}$  sob o esquema de perturbação na variável resposta.

De acordo com a análise realizada foi possível observar que o indivíduo 1 se destacou como possível observação influente sob alguns esquemas de perturbação em ambos os modelos paramétrico e semiparamétrico. Ele foi identificado também como um valor extremo no *boxplot* apresentado na Figura 5.2. Além disso, com relação aos ajustes obtidos para esse indivíduo, em ambos os casos os modelos tendem a subestimar as curvas individuais. Os indivíduos 2 e 12 se destacaram como possíveis observações influentes somente sob o esquema de perturbação na variável resposta em todas as direções de influência, mas os modelos propostos se ajustaram bem aos dados desses indivíduos. Já o indivíduo 5 ficou acima do ponto de corte sob o esquema de perturbação na variável resposta na direção de maior influência e com relação à distancia de Mahalanobis. Como dito anteriormente, esse indivíduo possui um valor alto na quarta medida, que é subestimada em ambos os modelos.

No capítulo seguinte é realizada uma discussão sobre o estudo desenvolvido nesta dissertação e são apontadas algumas propostas para trabalhos futuros.

# Capítulo 6

## Discussão e propostas futuras

Neste trabalho foi proposto o modelo misto semiparamétrico parcialmente não linear que pode ser uma alternativa para se estudarem dados correlacionados com estrutura não linear. Esse modelo foi apresentado como uma generalização de vários outros como, por exemplo, modelos mistos não lineares, modelos mistos lineares e modelos lineares mistos semiparamétricos. Foi discutido o processo de estimação dos parâmetros baseado na maximização do logaritmo da função de verossimilhança penalizada. O critério de informação AIC foi adotado para selecionar o valor ótimo do parâmetro de suavização. O método de influência local para o modelo proposto foi desenvolvido. Por fim, foi apresentada a aplicação realizada aos dados de concentração de *theophylline*, que são caracterizados pela relação não linear entre a variável resposta e os parâmetros de efeitos fixos e pela estrutura de correlação intraunidade experimental.

A aplicação foi norteadada pela comparação dos resultados entre os modelos paramétrico e semiparamétrico. Nesse desenvolvimento foi possível verificar que as curvas ajustadas são muito semelhantes graficamente, apesar de possuírem estimativas dos parâmetros com valores diferentes. Com relação ao estudo de influência local, os resultados indicaram que a capacidade para acomodar observações aberrantes varia de acordo com o esquema de perturbação, não sendo possível concluir qual modelo é melhor nesse sentido.

Durante o desenvolvimento deste trabalho surgiram alguns desafios computacionais. Uma das dificuldades em trabalhar com modelos não lineares é com relação aos valores iniciais para o algoritmo utilizado. Quando se consideram modelos não lineares semiparamétricos, essas dificuldades ampliam se consideravelmente. O algoritmo proposto exige valores iniciais adequados para se atingir a convergência. Além disso, apresentou algumas dificuldades de convergência para alguns valores do parâmetro de suavização.

Em conclusão, existem diversas possibilidades de continuação deste trabalho. Uma delas seria melhorar o processo de estimação ao considerar outros algoritmos. Além disso, é necessário realizar uma discussão com relação à existência e unicidade do estimador de máxima verossimilhança penalizada. Ademais, é possível considerar distribuições

com caudas mais pesadas do que a distribuição normal. Verifica-se também a importância de se realizar um estudo de simulação para verificar empiricamente as propriedades do modelo ajustado. Considera-se que seria relevante aplicar o MMSPNL a diferentes conjuntos de dados para avaliar sua performance em diferentes situações. Para finalizar, uma conclusão importante que se depreende da análise desenvolvida para o conjunto de dados de concentração de *theophylline* é que os pressupostos do modelo paramétrico estão bem definidos. Especificamente, a função paramétrica não linear descreve satisfatoriamente o mecanismo gerador dos dados. Caso contrário, o modelo semiparamétrico iria melhorar consideravelmente a modelagem estatística, ao fornecer maior flexibilidade à forma funcional que subjaz aos dados. Para verificar essa afirmação, pode-se considerar uma função não linear insatisfatória para descrever os dados e analisar o desempenho dos modelos paramétrico e semiparamétrico.

# Referências Bibliográficas

- Cook, D. (1986), ‘Assessment of local influence’, *Journal of the Royal Statistical Society - Series B* **48**, 133–169.
- Davidian, M. & Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman and Hall.
- De Boor, C. (1978), *A practical guide to splines*, Vol. 27, Springer-Verlag New York.
- do Carmo, M. P. (1976), *Differential geometry of curves and surfaces*, Vol. 2, Prentice-Hall Englewood Cliffs.
- Eilers, P. H. & Marx, B. D. (1996), ‘Flexible smoothing with b-splines and penalties’, *Statistical science* **11**, 89–102.
- Elmi, A., Ratcliffe, S. J., Parry, S. & Guo, W. (2011), ‘A b-spline based semiparametric nonlinear mixed effects model’, *Journal of Computational and Graphical Statistics* **20**, 492–509.
- Goodd, I. & Gaskins, R. (1971), ‘Nonparametric roughness penalties for probability densities’, *Biometrika* **58**, 255–277.
- Green, P. J. (1987), ‘Penalized likelihood for general semi-parametric regression models’, *International Statistical Review/Revue Internationale de Statistique* **55**, 245–259.
- Green, P. J. & Silverman, B. W. (1994), *Nonparametric regression and generalized linear models: a roughness penalty approach*, Chapman & Hall London.
- Ibacache-Pulgar, G., Paula, G. A. & Galea, M. (2012), ‘Influence diagnostics for elliptical semiparametric mixed models’, *Statistical Modelling* **12**, 165–193.
- Ke, C. & Wang, Y. (2001), ‘Semiparametric nonlinear mixed-effects models and their applications’, *Journal of the American Statistical Association* **96**, 1272–1298.
- Laird, N. M. & Ware, J. H. (1982), ‘Random-effects models for longitudinal data’, *Biometrics* **38**, 963–974.

- Lee, S. Y. & Xu, L. (2004), ‘Influence analyses of nonlinear mixed-effects models’, *Computational Statistics & Data Analysis* **45**, 321–341.
- Lindstrom, M. J. & Bates, D. M. (1990), ‘Nonlinear mixed effects models for repeated measures data’, *Biometrics* **46**, 673–687.
- Liu, W. & Wu, L. (2008), ‘A semiparametric nonlinear mixed-effects model with non-ignorable missing data and measurement errors for hiv viral data’, *Computational Statistics & Data Analysis* **53**, 112–122.
- Machado, R. J. M. & Russo, C. M. (2013), ‘Semiparametric partially nonlinear mixed-effects models with p-splines’, *In: Proceedings of the 28th International Workshop on Statistical Modelling, Muggeo, V. M. R., Capursi, V., Boscaino, G., Lovison, G. (Eds). Palermo: Università Degli Studi di Palermo* **1**, 251–256.
- Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, New York: Wiley.
- O’sullivan, F., Yandell, B. S. & Raynor Jr, W. J. (1986), ‘Automatic smoothing of regression functions in generalized linear models’, *Journal of the American Statistical Association* **81**, 96–103.
- Osório, F., Paula, G. A. & Galea, M. (2007), ‘Assessment of local influence in elliptical linear models with longitudinal structure’, *Computational Statistics & Data Analysis* **51**, 4354–4368.
- Pinheiro, J. C. & Bates, D. M. (2000), *Mixed-Effects Models in S and S-Plus*, New York:Springer.
- Pinheiro, J. C. & Bates, D. M. (2002), *Mixed-Effects Models in S and S-Plus*, Springer.
- Poon, W. Y. & Poon, Y. S. (1999), ‘Conformal normal curvature and assessment of local influence’, *Journal of the Royal Statistical Society - Series B* **61**, 51–61.
- Racine, J. S. (2012), ‘A primer on regression splines’, *CRAN. R-Project. Available online: [http://cran.rproject.org/web/packages/crs/vignettes/spline\\_primer.pdf](http://cran.rproject.org/web/packages/crs/vignettes/spline_primer.pdf)*.
- Reinsch, C. H. (1967), ‘Smoothing by spline functions’, *Numerische mathematik* **10**, 177–183.
- Russo, C. M. (2010), *Modelos não lineares elípticos para dados correlacionados*, Tese de Doutorado, IME - USP.
- Russo, C. M., Paula, G. A. & Aoki, R. (2009), ‘Influence diagnostics in nonlinear mixed-effects elliptical models’, *Computational Statistics & Data Analysis* **53**, 4143–4156.

- Salgado, F. A. O. (2006), *Diagnósticos de Influência em Modelos Elípticos com Efeitos Mistos*, Tese de Doutorado, IME - USP.
- Shen, X. (1997), ‘On methods of sieves and penalization’, *The Annals of Statistics* **25**, 2555–2591.
- Silverman, B. W. (1985), ‘Some aspects of the spline smoothing approach to non-parametric regression curve fitting’, *Journal of the Royal Statistical Society. Series B (Methodological)* **47**, 1–52.
- Vonesh, E. F. & Carter, R. L. (1992), ‘Mixed-effects nonlinear regression for unbalanced repeated measures’, *Biometrics* **48**, 1–17.
- Wang, Y. (1998), ‘Mixed effects smoothing spline analysis of variance’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**(1), 159–174.
- Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998), ‘Semiparametric stochastic mixed models for longitudinal data’, *Journal of the American Statistical Association* **93**, 710–719.

# Apêndice A

## Dados de concentração de theophylline

A Tabela A.1 representa os dados de concentração de Theophylline. A concentração de theophylline é representada pela variável conc.

Tabela A.1: Dados de concentração de theophylline

	Sujeito	Peso	Dose	Tempo	conc
1	1	79.60	4.02	0.00	0.74
2	1	79.60	4.02	0.25	2.84
3	1	79.60	4.02	0.57	6.57
4	1	79.60	4.02	1.12	10.50
5	1	79.60	4.02	2.02	9.66
6	1	79.60	4.02	3.82	8.58
7	1	79.60	4.02	5.10	8.36
8	1	79.60	4.02	7.03	7.47
9	1	79.60	4.02	9.05	6.89
10	1	79.60	4.02	12.12	5.94
11	1	79.60	4.02	24.37	3.28
12	2	72.40	4.40	0.00	0.00
13	2	72.40	4.40	0.27	1.72
14	2	72.40	4.40	0.52	7.91
15	2	72.40	4.40	1.00	8.31
16	2	72.40	4.40	1.92	8.33
17	2	72.40	4.40	3.50	6.85
18	2	72.40	4.40	5.02	6.08
19	2	72.40	4.40	7.03	5.40
20	2	72.40	4.40	9.00	4.55
21	2	72.40	4.40	12.00	3.01
22	2	72.40	4.40	24.30	0.90

	Sujeito	Peso	Dose	Tempo	conc
23	3	70.50	4.53	0.00	0.00
24	3	70.50	4.53	0.27	4.40
25	3	70.50	4.53	0.58	6.90
26	3	70.50	4.53	1.02	8.20
27	3	70.50	4.53	2.02	7.80
28	3	70.50	4.53	3.62	7.50
29	3	70.50	4.53	5.08	6.20
30	3	70.50	4.53	7.07	5.30
31	3	70.50	4.53	9.00	4.90
32	3	70.50	4.53	12.15	3.70
33	3	70.50	4.53	24.17	1.05
34	4	72.70	4.40	0.00	0.00
35	4	72.70	4.40	0.35	1.89
36	4	72.70	4.40	0.60	4.60
37	4	72.70	4.40	1.07	8.60
38	4	72.70	4.40	2.13	8.38
39	4	72.70	4.40	3.50	7.54
40	4	72.70	4.40	5.02	6.88
41	4	72.70	4.40	7.02	5.78
42	4	72.70	4.40	9.02	5.33
43	4	72.70	4.40	11.98	4.19
44	4	72.70	4.40	24.65	1.15
45	5	54.60	5.86	0.00	0.00
46	5	54.60	5.86	0.30	2.02
47	5	54.60	5.86	0.52	5.63
48	5	54.60	5.86	1.00	11.40
49	5	54.60	5.86	2.02	9.33
50	5	54.60	5.86	3.50	8.74
51	5	54.60	5.86	5.02	7.56
52	5	54.60	5.86	7.02	7.09
53	5	54.60	5.86	9.10	5.90
54	5	54.60	5.86	12.00	4.37
55	5	54.60	5.86	24.35	1.57



	Sujeito	Peso	Dose	Tempo	conc
56	6	80.00	4.00	0.00	0.00
57	6	80.00	4.00	0.27	1.29
58	6	80.00	4.00	0.58	3.08
59	6	80.00	4.00	1.15	6.44
60	6	80.00	4.00	2.03	6.32
61	6	80.00	4.00	3.57	5.53
62	6	80.00	4.00	5.00	4.94
63	6	80.00	4.00	7.00	4.02
64	6	80.00	4.00	9.22	3.46
65	6	80.00	4.00	12.10	2.78
66	6	80.00	4.00	23.85	0.92
67	7	64.60	4.95	0.00	0.15
68	7	64.60	4.95	0.25	0.85
69	7	64.60	4.95	0.50	2.35
70	7	64.60	4.95	1.02	5.02
71	7	64.60	4.95	2.02	6.58
72	7	64.60	4.95	3.48	7.09
73	7	64.60	4.95	5.00	6.66
74	7	64.60	4.95	6.98	5.25
75	7	64.60	4.95	9.00	4.39
76	7	64.60	4.95	12.05	3.53
77	7	64.60	4.95	24.22	1.15
78	8	70.50	4.53	0.00	0.00
79	8	70.50	4.53	0.25	3.05
80	8	70.50	4.53	0.52	3.05
81	8	70.50	4.53	0.98	7.31
82	8	70.50	4.53	2.02	7.56
83	8	70.50	4.53	3.53	6.59
84	8	70.50	4.53	5.05	5.88
85	8	70.50	4.53	7.15	4.73
86	8	70.50	4.53	9.07	4.57
87	8	70.50	4.53	12.10	3.00
88	8	70.50	4.53	24.12	1.25

	Sujeito	Peso	Dose	Tempo	conc
89	9	86.40	3.10	0.00	0.00
90	9	86.40	3.10	0.30	7.37
91	9	86.40	3.10	0.63	9.03
92	9	86.40	3.10	1.05	7.14
93	9	86.40	3.10	2.02	6.33
94	9	86.40	3.10	3.53	5.66
95	9	86.40	3.10	5.02	5.67
96	9	86.40	3.10	7.17	4.24
97	9	86.40	3.10	8.80	4.11
98	9	86.40	3.10	11.60	3.16
99	9	86.40	3.10	24.43	1.12
100	10	58.20	5.50	0.00	0.24
101	10	58.20	5.50	0.37	2.89
102	10	58.20	5.50	0.77	5.22
103	10	58.20	5.50	1.02	6.41
104	10	58.20	5.50	2.05	7.83
105	10	58.20	5.50	3.55	10.21
106	10	58.20	5.50	5.05	9.18
107	10	58.20	5.50	7.08	8.02
108	10	58.20	5.50	9.38	7.14
109	10	58.20	5.50	12.10	5.68
110	10	58.20	5.50	23.70	2.42
111	11	65.00	4.92	0.00	0.00
112	11	65.00	4.92	0.25	4.86
113	11	65.00	4.92	0.50	7.24
114	11	65.00	4.92	0.98	8.00
115	11	65.00	4.92	1.98	6.81
116	11	65.00	4.92	3.60	5.87
117	11	65.00	4.92	5.02	5.22
118	11	65.00	4.92	7.03	4.45
119	11	65.00	4.92	9.03	3.62
120	11	65.00	4.92	12.12	2.69
121	11	65.00	4.92	24.08	0.86
122	12	60.50	5.30	0.00	0.00
123	12	60.50	5.30	0.25	1.25
124	12	60.50	5.30	0.50	3.96
125	12	60.50	5.30	1.00	7.82
126	12	60.50	5.30	2.00	9.72
127	12	60.50	5.30	3.52	9.75
128	12	60.50	5.30	5.07	8.57
129	12	60.50	5.30	7.07	6.59
130	12	60.50	5.30	9.03	6.11
131	12	60.50	5.30	12.05	4.57
132	12	60.50	5.30	24.15	1.17

# Apêndice B

## $B$ -splines e $P$ -splines

Neste apêndice serão apresentadas mais informações sobre  $B$ -splines, a fim de proporcionar um entendimento mais amplo sobre os termos não paramétricos no modelo 3.1 <sup>1</sup>. Considere  $N + 2$  valores reais  $t_i$ , chamados nós

$$t_0 \leq t_1 \leq \dots \leq t_{N+1}.$$

Os  $N \geq 0$  valores  $t_i$  são chamados nós internos e sempre existem dois pontos extremos,  $t_0$  e  $t_{N+1}$ .

Seja  $\mathbf{t} = \{t_i; i \in \mathbb{Z}\}$  uma sequência de números reais não decrescentes ( $t_i \leq t_{i+1}$ ). Defina-se uma sequência de nós por

$$t_{-(m-1)} = \dots = t_0 \leq t_1 \leq \dots \leq t_N \leq t_{N+1} = \dots = t_{N+m},$$

em que os nós extremos foram incluídos  $n = m - 1$  vezes (necessário para a construção recursiva das  $B$ -splines). Assim, pode-se indexar o conjunto de nós por  $i = 0, \dots, N + 2m - 1$ .

Dada uma sequência de nós  $t_i$ ,  $i = 0, \dots, N + 2m - 1$ , a  $i$ -ésima função base  $B$ -spline de ordem  $j$ ,  $B_{i,j}$ , é definida pela relação de recorrência (De Boor 1978, cap. 9)

$$B_{i,0}(x) = \begin{cases} 1, & t_i \leq x < t_{i+1} \\ 0, & \text{caso contrário} \end{cases} \quad (\text{B.1})$$

$$B_{i,j+1}(x) = \alpha_{i,j+1}(x)B_{i,j}(x) + [1 - \alpha_{i+j,j+1}(x)]B_{i+1,j}(x), \quad (\text{B.2})$$

em que

$$\alpha_{i,j}(x) = \begin{cases} \frac{x - t_i}{t_{i+j} - t_i}, & t_{i+j} \neq t_i \\ 0, & \text{caso contrário} \end{cases} \quad (\text{B.3})$$

---

<sup>1</sup>A descrição foi baseada em Racine (2012).

Para realizar os cálculos acima,  $\frac{0}{0}$  é definido como 0.

Uma *B-spline* de grau  $n$  (de ordem  $m = n + 1$ ) é uma curva paramétrica composta por uma combinação linear de bases de *B-splines*  $B_{i,n}(x)$  de grau  $n$ , dada por

$$B(x) = \sum_{i=0}^{N+n} \beta_i B_{i,n}(x), \quad (\text{B.4})$$

em que  $x \in [t_0, t_{N+1}]$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{N+n})^\top$  é um vetor de números reais.

Como exemplo ilustrativo de uma função *B-spline*, considere um conjunto com três nós interiores ( $N = 3$ ), dados por  $\{0.25, 0.5, 0.75\}$ , em que os nós extremos são  $\{0, 1\}$ . O grau da *spline* será  $n = 3$ . Portanto, sua ordem será  $m = 4$ . O conjunto de todos os nós necessários para construir a *B-spline* é

$$\{0, 0, 0, 0, 0.25, 0.5, 0.75, 1, 1, 1, 1\}$$

e o número de funções base é  $K = N + m = 7$ . As sete funções base de *B-spline* serão denotadas por  $B_{0,3}, \dots, B_{6,3}$ . A Figura B.1 representa esse exemplo de uma *B-spline* de terceiro grau e três nós equidistantes.

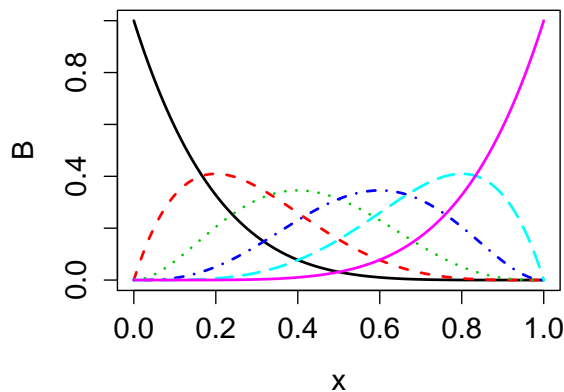


Figura B.1: Funções base de *B-splines* de terceiro grau e 3 nós interiores.

O método de suavização com *P-splines* consiste em aproximar uma função não paramétrica com *B-splines* e considerar a função de penalidade baseada no operador de diferenças (de coeficientes de *B-splines* adjacentes)

$$J(f) = \sum_{j=k+1}^{N+n} (\Delta^k \beta_j)^2, \quad (\text{B.5})$$

em que  $f$  é a função a ser estimada em um problema de regressão não paramétrica,  $k$  é a ordem das diferenças,  $\Delta^k \beta_j = \Delta \Delta^{k-1} \beta_j$  e  $\Delta \beta_j = \beta_j - \beta_{j-1}$  (Eilers & Marx 1996).

Para calcular a matriz  $\mathbf{B}$  de representação de *B-splines* para um dado vetor  $\mathbf{x}$  no  $\mathbb{R}$ , baseando-se em *B-spline* de grau `bdeg`, pode-se usar a função

```
bspline<-function(x, xl, xr, ndx, bdeg) {  
  dx <- (xr-xl) /ndx  
  knots <- seq(xl-bdeg*dx,xr+bdeg*dx, by=dx)  
  B<-spline.des(knots, x, bdeg+ 1, 0*x)$design  
  B  
}
```

em que `xl` e `xr` são o mínimo e o máximo do domínio de  $\mathbf{x}$ , respectivamente, e existem `ndx` intervalos no domínio de  $\mathbf{x}$ . A matriz  $\mathbf{D}_k$  de representação do operador de diferenças pode ser calculada usando-se o código

```
D <- diag(ncol(B))  
  for (k in 1:pord) D <- diff(D)
```

em que `pord` representa a ordem das diferenças.