Variedades de Brieskorn e Polinômios Quase-homogêneos

Sadao Massago

Orientador: Prof. Dr. *Washington Luiz Marar*

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos, Universidade de São Paulo, como parte do requisitos para obtenção do título de Mestre em Ciências - Área: Matemática

São Carlos, fevereiro de 1995.

*Dedico este trabalho
à minha família
e ao Luis Henrique.*

Agradecimentos

A conclusão desta dissertação não seria possível sem o auxílio de algumas pessoas e instituições que, de alguma forma, contribuíram para este projeto, às quais expresso meus sinceros agradecimentos.

Ao Prof. Dr. Jorge Alberto Achcar pela orientação sempre dedicada, pela participação ativa no estudo e pelo incentivo em todo o decorrer da pesquisa.

Ao Prof. Dr. Josemar Rodrigues pelos comentários e valiosas sugestões quando da realização do meu exame de qualificação.

Ao Prof. Dr. Gilberto Alvarenga Paula e Prof. Dr. Carlos Alberto Diniz pelos comentários e sugestões para a versão final desta dissertação.

À CAPES pelo apoio financeiro.

Aos funcionários, professores, colegas de pós-graduação pelo excelente convívio e a todos aqueles que de alguma forma contribuíram para a realização deste trabalho e especialmente aos amigos, Carlos Aparecido, Elisa, Gaby, José Carlos, Josmar, Masé, Sadao, Silvana, Sílvia, Vera, Walmir, pela amizade, carinho e apoio durante este período, dentro e fora do ICMSC.

Em especial ao Luis Henrique pelo carinho, paciência e incentivo.

Resumo

Apresentamos neste trabalho, um estudo sobre modelos não-lineares com ênfase num modelo muito utilizado em aplicações biológicas, o modelo de Michaelis-Menten. Em especial, destacamos a importância de uma boa parametrização que garanta boas inferências assintóticas, e introduzimos uma análise Bayesiana aproximada usando densidades a priori não-informativas e informativas. Nessas aproximações, enfatizamos o uso do método de Laplace para obter os sumários a posteriori de interesse. Também consideramos o uso de distribuições não-normais para o erro e o uso de técnicas de transformações para os dados sob o enfoque Bayesiano.

Finalizamos ilustrando com vários exemplos de aplicação considerando o modelo de Michaelis-Menten.

Abstract

In this work, we present a study on non-linear models with special consideration for a model very popular in biological applications, the Michaelis-Menten model. In special, we discuss the importance of a good parametrization which implies in good asymptotical inferences, and we introduce a Bayesian analysis using approximation methods considering noninformative and informative prior densities. In these approximations, we explore the use of Laplace's method to get the posterior summaries of interest. We also consider the use of non-normal distributions for the error and the use of transformation techniques for the data considering the Bayesian approach.

We illustrate with some examples.

Terminologia e Notação

- As matrizes e os vetores são denotados por letras em negrito $\mathbf{A}$ e $\mathbf{a}$, respectivamente, e escalares por letras itálicas.

- Se $\mathbf{a}$ é um vetor ($n \times 1$) com elementos $a_1, a_2, \dots, a_n$, podemos escrever $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, e a norma ou comprimento de $\mathbf{a}$ é denotada por $\|\mathbf{a}\|$. Assim,

$$\|\mathbf{a}\| = (\mathbf{a}'\mathbf{a})^{1/2} = (a_1^2 + a_2^2 + \dots + a_n^2)^{1/2}$$

- $\mathbb{R}^d$ representa o espaço Euclidiano d -dimensional.
- $R[\mathbf{A}]$ é chamado espaço coluna de $\mathbf{A}$.
- Se x e y são variáveis aleatórias, então os símbolos $E(x)$, $Var(x)$, $cov(x, y)$ e $E(x/y)$ representam a esperança, variância, covariância e esperança condicional, respectivamente. Analogamente, para o caso multivariado, $\mathcal{E}(\mathbf{x})$ representa o valor esperado do vetor aleatório $\mathbf{x}$, $\mathcal{C}[\mathbf{x}, \mathbf{y}] = \begin{bmatrix} cov[x_i, y_j] \end{bmatrix} = \mathcal{C}[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)']$ representa a covariância de dois vetores aleatórios $\mathbf{x}$ e $\mathbf{y}$ com médias $\boldsymbol{\mu}_x$ e $\boldsymbol{\mu}_y$, respectivamente, e $\mathcal{V}[\mathbf{x}] = \mathcal{C}[\mathbf{x}, \mathbf{x}]$ é a matriz de variância e covariância de $\mathbf{x}$.

- Em geral, constantes são representadas por letras minúsculas.

- As funções de verossimilhança dos parâmetros envolvidos são representadas por $L(\cdot)$, o logaritmo da função de verossimilhança por $l(\cdot)$, as densidades a priori dos parâmetros envolvidos por $p(\cdot)$, as densidades a posteriori por $p(\cdot / \text{dados})$ e a densidade preditiva de uma observação futura por $f(\cdot / \text{dados})$.

- Letras gregas iniciam os parâmetros dos modelos.

Índice

Capítulo 1

INTRODUÇÃO AOS MODELOS NÃO-LINEARES	01
1.1.Objetivos Principais da Pesquisa.....	02
1.2.Apresentação dos Capítulos	03
1.3.Modelos Não-Lineares	04
1.4.Estimação de Parâmetros em Modelos Não-Lineares	06
1.4.1.Mínimos Quadrados Não-Lineares.....	06
1.4.2.Aproximação Linear	08
1.4.3.Métodos Numéricos.....	10
1.4.4.Estimação de Máxima Verossimilhança.....	13
1.4.4.1.Erros Normais	13
1.5.Contornos de Confiança.....	15
1.6.Comparação de Resultados Assintóticos para Modelos Não-Lineares com Resultados Exatos para Modelos Lineares	16
1.6.1.Modelos Lineares	16
1.6.2.Modelos Não-Lineares.....	17
1.7.Problemas Práticos na Estimação de Modelos Não-Lineares	18
1.7.1.Convergência dos Métodos Iterativos	18
1.7.2.Validade da Inferência Assintótica.....	20
1.7.2.1.Regões de Confiança.....	20
1.7.2.2.Efeitos de Curvatura.....	21
1.7.2.2.1.Representação Geométrica dos Modelos de Regressão Não- Lineares no Espaço Amostral.....	22
1.7.2.2.2.Medidas de Curvatura de Bates e Watts	27
1.7.2.2.3.Medida do Vício de Box	31

Capítulo 2

INFERÊNCIA CLÁSSICA.....34

2.1.Intervalos de Confiança Assintóticos.....	34
2.2.Intervalos de Confiança para Funções Não-Lineares de θ	35
2.3.Intervalos de Confiança para Previsão.....	36
2.4.Regões de Confiança.....	37
2.5.Métodos Assintóticos baseados na Função de Verossimilhança	38
2.6.Testes de Hipóteses.....	39
2.6.1.Hipóteses Lineares.....	39
2.7.Análise Clássica para o Modelo de Michaelis-Menten baseada na Normalidade Assintótica	42
2.7.1.Formulação do Modelo.....	42
2.7.2.A Função de Verossimilhança	43
2.7.3.Estimadores de Máxima Verossimilhança.....	44
2.7.3.1.Inferências sobre os Parâmetros do Modelo.....	45

Capítulo 3

ALGUNS ASPECTOS DE REPARAMETRIZAÇÃO.....47

3.1.Algumas Reparametrizações para o Modelo de Michaelis-Menten.....	47
3.1.1.Reparametrização proposta por Bates e Watts.....	48
3.1.2.Reparametrização proposta por Seber e Wild.....	48
3.2.Medidas e Diagnósticos para a Normalidade.....	49
3.2.1.T-Plot de Hills e Smith	49
3.2.2.Terceira Derivada Padronizada do Logaritmo da Verossimilhança.....	51
3.3.Inferências para o Modelo Michaelis-Menten considerando as Reparametrizações Propostas	52
3.4.Inferências para o Modelo de Michaelis-Menten considerando a Função de Verossimilhança <i>Profile</i>	56
3.4.1.Reparametrização proposta por Bates e Watts para a Função de Verossimilhança <i>Profile</i>	58
3.4.2.Reparametrização proposta por Seber e Wild para a Função de Verossimilhança <i>Profile</i>	60

Capítulo 4

ESTIMAÇÃO BAYESIANA PARA MODELOS NÃO-LINEARES.62

4.1.Fórmula de Bayes	62
4.2.Densidades a Priori para os Parâmetros.....	63
4.3.Densidades a Posteriori para os Parâmetros.....	65
4.3.1.Momentos a Posteriori.....	66
4.3.2.Densidades Preditivas	67
4.4.Análise Bayesiana do Modelo de Michaelis-Menten considerando uma Densidade a Priori Não-Informativa	68
4.4.1.Formulação do Modelo.....	68
4.4.2.Uma Densidade a Priori Não-Informativa para θ e σ	68
4.4.3.Densidade a Posteriori Conjunta para θ e σ	69
4.4.3.1.Densidades a Posteriori Marginais para θ_1 e θ_2	70
4.4.4.Momentos a Posteriori para θ_1 e θ_2	73
4.4.5.Densidade Preditiva para uma Observação Futura	74
4.5.Análise Bayesiana do Modelo de Michaelis-Menten considerando uma Densidade a Priori Informativa	75
4.5.1.Uma Densidade a Priori Informativa para θ_1 , θ_2 e σ	75
4.5.2.Densidade a Posteriori Conjunta para θ_1 , θ_2 e σ	77
4.5.3.Densidades a Posteriori Marginais para θ_1 e θ_2	77
4.5.4.Densidade Preditiva para uma Observação Futura	79
4.6.Reparametrizações para o Modelo de Michaelis-Menten.....	81

Capítulo 5

MODELO DE MICHAELIS-MENTEN SEM A SUPOSIÇÃO DE NORMALIDADE PARA O ERRO.86

5.1.A Escolha da Classe de Distribuições Potência Exponencial	86
5.2.Critério de Robustez e Inferência Robusta	89
5.3.Análise Bayesiana do Modelo de Michaelis-Menten considerando β conhecido	90
5.3.1.Densidade a Priori Não-Informativa para θ_1 , θ_2 e σ	90
5.3.2.Densidade a Posteriori Conjunta para θ_1 , θ_2 e σ	91
5.3.3.Densidade a Posteriori Conjunta para θ_1 e θ_2	91

5.4. Análise Bayesiana do Modelo de Michaelis-Menten considerando β desconhecido	92
5.4.1. Densidade a Priori Não-Informativa para θ_1 , θ_2 , σ e β	93
5.4.2. Densidade a Posteriori Conjunta para θ_1 , θ_2 , σ e β	93
5.4.3. Densidade a Posteriori Conjunta para θ_1 , θ_2 e β	94
5.4.4. Densidade a Posteriori Marginal para β	94

Capítulo 6

TRANSFORMAÇÕES DOS DADOS.	98
6.1. Uma Família Paramétrica de Transformações de Potência.....	99
6.2. Transformação da Variável Dependente.....	100
6.3. Análise Bayesiana para o Modelo de Michaelis-Menten	101
6.3.1. Densidade a Priori para λ , θ e σ	101
6.3.2. Densidade a Posteriori Conjunta para λ , θ e σ	102
6.3.3. Densidade a Posteriori Marginal para λ	102

Capítulo 7

APLICAÇÕES.	104
7.1. Exemplo considerando o conjunto de dados introduzidos por Treloar, (1974), e o modelo de Michaelis-Menten.	104
7.2. Exemplo considerando um conjunto de dados introduzido por Bates e Watts, (1980), e o modelo de Michaelis-Menten.	123
7.3. Exemplo em que os dados foram gerados do modelo de Michaelis-Menten, tal que os erros tenham distribuição lognormal.	139
7.4. Exemplo considerando dois conjuntos de dados para estudar o efeito de não-normalidade nas inferências do modelo de Michaelis-Menten.....	148

Capítulo 8

CONCLUSÕES FINAIS E CONSIDERAÇÕES FUTURAS.	167
---	-----

Apêndice A

Teoria Assintótica	168
--------------------------	-----

A.1. Estimadores de Mínimos Quadrados	168
---	-----

A.1.1. Existência dos Estimadores de Mínimos Quadrados.....	168
---	-----

A.1.2. Consistência	169
---------------------------	-----

A.1.3. Normalidade Assintótica.....	170
-------------------------------------	-----

Apêndice B

Decomposição QR de uma Matriz	174
---------------------------------------	-----

Apêndice C

Estudo de Simulação.	177
---------------------------	-----

Apêndice D

Alguns Métodos de Resolução de Integrais em Inferência Bayesiana.....	181
---	-----

D.1. Aproximação de Tierney e Kadane.....	181
---	-----

D.2. Métodos Numéricos	183
------------------------------	-----

D.2.1. Regra de Simpson	183
-------------------------------	-----

D.2.2. Regra Gaussiana	184
------------------------------	-----

Apêndice E

E.1. Integral Gama Invertida.....	186
-----------------------------------	-----

E.2. Expressões utilizadas para a obtenção de Densidades a Posteriori Marginais de θ	186
---	-----

E.3. Expressões utilizadas para a obtenção de Momentos a Posteriori para θ	187
---	-----

E.3.1. Média a Posteriori para θ_1	187
---	-----

E.3.2. Média a Posteriori para θ_2	189
---	-----

E.4. Expressões utilizadas na obtenção de Densidade Preditiva para uma Observação Futura	190
--	-----

E.5.Expressões utilizadas na obtenção de Densidade a Posteriori Marginais para	
β	192
E.5.1.Densidade a Posteriori Marginal para β_1	192
E.5.2.Densidade a Posteriori Marginal para β_2	192
 Apêndice F	
Alguns Programas desenvolvidos.....	193
 Referências e Bibliografia Consultada.	218

Capítulo 1

INTRODUÇÃO AOS MODELOS NÃO-LINEARES

Os modelos lineares representam uma ampla variedade de relações existentes entre as variáveis de interesse para os pesquisadores, mas existem alguns casos, nos quais, o modelo linear não é apropriado. Quando as informações teóricas em relação ao modelo nos levam a um modelo não-linear, estudos recentes revelam que devemos, geralmente, preferir ajustar tal modelo, sempre que possível, do que tentar ajustar um modelo linear talvez menos realístico.

Em várias áreas de aplicação, como por exemplo, em engenharia química, biologia, física e química, os modelos não-lineares são muito utilizados e a análise estatística desses modelos é de grande interesse prático. A análise estatística desses modelos pode oferecer grandes dificuldades, pois os algoritmos existentes podem não convergir e os resultados inferenciais baseados numa aproximação linear podem não ser confiáveis.

Em regressão não-linear, o problema de estimação depende essencialmente da utilização de métodos clássicos de otimização não-linear que foram introduzidos por Newton, Cauchy e Gauss. Recentemente, houve um grande progresso no desenvolvimento dessas técnicas e existe uma vasta literatura sobre os mais variados métodos para os mais variados tipos de problemas, inclusive o ajuste de dados. Segundo Ratkowsky (1983), apesar da grande variedade de métodos para a obtenção das estimativas de mínimos quadrados, o estudo das propriedades desses estimadores ainda encontra-se em fase inicial e recentemente tem surgido alguns resultados importantes, mas em número limitado.

Usando a teoria clássica assintótica, apresentaremos a importância de boas parametrizações com o intuito de obter boa normalidade da função de verossimilhança, que implicam geralmente em boa precisão aos resultados assintóticos.

Considerando o modelo não-linear de Michaelis-Menten (Bates e Watts, 1988), discutiremos alguns aspectos sobre a precisão das inferências para os parâmetros do modelo, as quais são obtidas por métodos assintóticos e como sabemos, a precisão desses resultados usualmente dependem do conjunto de dados e de uma parametrização apropriada.

Assim, uma alternativa para a obtenção de inferências precisas para os parâmetros é o uso de métodos Bayesianos. Neste trabalho, damos ênfase à utilização de técnicas Bayesianas para os modelos não-lineares, visando maior eficiência e precisão nos resultados aproximados.

Na prática, usualmente, os pesquisadores têm interesse em obter inferências sobre os parâmetros originais, desde que esses tenham interpretações físicas importantes. Dependendo da parametrização utilizada, as inferências para os parâmetros originais podem ser imprecisas, assim, uma alternativa apropriada para obter precisão nas inferências desses parâmetros é através de métodos Bayesianos.

1.1. Objetivos Principais da Pesquisa

O projeto de pesquisa apresentado neste trabalho tem como objetivos fazer uma revisão dos resultados recentes para a análise estatística dos modelos não-lineares e um estudo comparativo a partir de uma análise Bayesiana, considerando alguns casos especiais.

Em particular destacamos,

- i)* A construção de modelos não-lineares;
- ii)* Os métodos clássicos de estimação e inferência estatística;
- iii)* As medidas de curvatura e suas implicações nas inferências obtidas;
- iv)* Considerando, densidades a priori não-informativas (ver por exemplo, Box e Tiao, 1973) e densidades a priori informativas para os parâmetros envolvidos, pretendemos determinar uma análise Bayesiana do modelo de Michaelis-Menten (ver Bates e Watts, 1988); explorando o uso do método de Laplace para aproximação de integrais (ver por exemplo, Tierney e Kadane, 1986) quando as densidades a posteriori de interesse não apresentam soluções analíticas;
- v)* Uso de métodos numéricos e de aproximação de integrais para a obtenção de momentos a posteriori e densidades preditivas de interesse;
- vi)* Pretendemos avaliar alguns aspectos de reparametrização na precisão de alguns resultados clássicos que são sensíveis a escolha de uma parametrização adequada.
- vii)* Considerando a distribuição potência exponencial para o erro do modelo de Michaelis-Menten, pretendemos verificar os possíveis efeitos nas inferências e verificar a robustez ou não das mesmas. Também, quando não dispomos de normalidade para o erro do modelo, consideraremos a família de transformação de potência (Box e Cox, 1964) para a variável dependente e assim, faremos inferências usando métodos Bayesianos aproximados.

1.2. Apresentação dos Capítulos

No primeiro capítulo, consideramos alguns dos métodos iterativos de estimação paramétrica mais conhecidos para a análise dos modelos não-lineares e estudamos algumas das medidas de não-linearidade mais recentes; as medidas de curvatura de Bates e Watts (1980) e a medida do vício proposta por Box (1971).

No capítulo 2, apresentamos, sob o enfoque clássico, as inferências estatísticas para os modelos não-lineares.

No capítulo 3, introduzimos algumas reparametrizações que melhorem a normalidade da função de verossimilhança. Exploramos o método gráfico baseado no *T-Plot* de Hills e Smith (1993) para avaliar as reparametrizações consideradas e utilizamos as medidas de normalidade introduzidas por Kass e Slate (1992), com a finalidade de assegurar a normalidade nas transformações obtidas.

No capítulo 4, considerando densidades a priori não-informativas de Jeffreys (ver Box e Tiao, 1973) e densidades a priori informativas para os parâmetros envolvidos, desenvolvemos uma análise Bayesiana para o modelo de Michaelis-Menten, assumindo uma distribuição normal para os erros.

Quando não temos normalidade para o erro no modelo de Michaelis-Menten, consideramos, no capítulo 5, a família de transformação de potência (Box e Cox, 1964) para a variável dependente do modelo e usando métodos Bayesianos obtemos inferências para os parâmetros de interesse.

No capítulo 6, verificamos os possíveis efeitos nas inferências do modelo de Michaelis-Menten, com erros não normais, considerando a distribuição potência exponencial (Box e Tiao, 1973) para o erro.

Exemplos de aplicação das metodologias, considerando o modelo de Michaelis-Menten com vários conjuntos de dados, são apresentados no capítulo 7 e finalmente no capítulo 8 são apresentadas as conclusões finais e comentários sobre perspectivas futuras.

Nas seções seguintes, apresentamos uma breve revisão dos resultados clássicos aproximados para os modelos não-lineares que têm sido recentemente muito explorados na literatura.

1.3. Modelos Não-Lineares

Um modelo de regressão é dito não-linear, se ele é não-linear como função dos seus parâmetros. Isto implica que as derivadas da função esperança com respeito aos parâmetros envolvem pelo menos um dos parâmetros.

Consideramos uma situação em que os dados consistem de uma resposta y que depende de p variáveis independentes x_i , e que pode ser representada pelo modelo de regressão não-linear,

$$y_i = f(x_i; \theta) + \varepsilon_i, \quad (i=1, 2, \dots, n) \quad (1.1)$$

ou na notação matricial,

$$Y = f(\theta) + \varepsilon \quad (1.2)$$

onde:

$Y' = (y_1, \dots, y_n)$ é o vetor ($n \times 1$) de observações ou de variáveis dependentes;

$f'(\theta) = (f(x_1; \theta), \dots, f(x_n; \theta))$ é o vetor ($n \times 1$) de funções de regressão ou função esperança;

$x'_i = (x_{i1}, \dots, x_{ip})$ são as variáveis independentes ou regressores;

$\theta' = (\theta_1, \dots, \theta_p)$ é o vetor ($p \times 1$) de parâmetros desconhecidos;

$\varepsilon' = (\varepsilon_1, \dots, \varepsilon_n)$ é o vetor ($n \times 1$) de erros experimentais (parte estocástica);

e as respostas esperadas são funções não-lineares dos parâmetros.

A seguir, vamos apresentar alguns exemplos de modelos não-lineares:

Exemplo 1.3.1. Consideramos um experimento destinado à física do calor, onde unidades de um certo metal foram aquecidas até 130°F e foram anotados os tempos e as temperaturas correspondentes quando deixadas em esfriamento. A temperatura ambiente durante o experimento foi de 60°F, isto é, sob a lei de Newton sobre esfriamento, que estabelece que $df/dt = -\theta(f - T_o)$, onde T_o é a temperatura ambiente, temos o modelo:

$$f(t, \theta) = 60 + 70e^{-\theta t} \quad (1.3)$$

Como $\partial f / \partial \theta = -70te^{-\theta t}$ depende do parâmetro θ , este modelo é não-linear.

Exemplo 1.3.2. O modelo de Michaelis-Menten para cinética de enzimas relaciona a velocidade inicial de uma reação com a concentração x de um substrato, através da equação,

$$f(x, \theta) = \frac{\theta_1 x}{\theta_2 + x} \quad (1.4)$$

Diferenciando f em relação a θ_1 e θ_2 , temos,

$$\frac{\partial f}{\partial \theta_1} = \frac{x}{\theta_2 + x} \quad (1.5)$$

$$\frac{\partial f}{\partial \theta_2} = \frac{-\theta_1 x}{(\theta_2 + x)^2}$$

e como as derivadas acima envolvem pelo menos um dos parâmetros, este modelo é não-linear.

Exemplo 1.3.3. (Modelos de Crescimento). Consideramos os modelos de crescimento não-lineares dados por,

$$\text{Gompertz: } y = \alpha \exp[-\exp(\beta - \gamma x)] \quad (1.6)$$

$$\text{Logístico: } y = \frac{\alpha}{1 + \exp(\beta - \gamma x)} \quad (1.7)$$

$$\text{Richards: } y = \frac{\alpha}{[1 + \exp(\beta - \gamma x)]^{1/\delta}} \quad (1.8)$$

$$\text{Morgan-Mercer-Flodin (MMF): } y = \frac{\beta\gamma + \alpha x^\delta}{\gamma + x^\delta} \quad (1.9)$$

Esses modelos são utilizados em muitas áreas de aplicação, como por exemplo: Biologia, Agricultura, Engenharia, Economia, etc. Maiores detalhes sobre os modelos de crescimento, ver por exemplo, Ratkowsky (1983).

1.4. Estimação de Parâmetros em Modelos Não-Lineares

O problema inicial é a obtenção de estimativas de θ à partir de um conjunto de dados observados.

1.4.1. Mínimos Quadrados Não-Lineares

Supor que temos n observações (x_i, y_i) , $i = 1, 2, \dots, n$, de um modelo de regressão não-linear, com uma relação funcional f conhecida, dada por,

$$y_i = f(x_i; \theta^*) + \varepsilon_i, \quad (i = 1, 2, \dots, n) \quad (1.10)$$

onde $E(\varepsilon_i) = 0$, x_i é um vetor $k \times 1$ e θ^* é o verdadeiro valor de θ , $\theta^* \in \Theta$, um subconjunto de R^p . O estimador de mínimos quadrados de θ^* , denotado por $\hat{\theta}$, minimiza a soma de quadrados residual dada por,

$$S(\theta) = \sum_{i=1}^n [y_i - f(x_i; \theta)]^2 \quad (1.11)$$

onde $\theta \in \Theta$.

Devemos ressaltar que $S(\theta)$ pode ter vários mínimos em adição ao mínimo absoluto $\hat{\theta}$. Assumindo que os erros (ε_i) são independentes e identicamente distribuídos com variância σ^2 , $\hat{\theta}$ e $s^2 = S(\hat{\theta})/(n-p)$ são estimadores consistentes de θ^* e σ^2 respectivamente. Sob condições de regularidade, ver Apêndice (A), $\hat{\theta}$ tem distribuição assintoticamente normal quando $n \rightarrow \infty$.

Quando cada $f(x_i; \theta)$ é diferenciável em relação a θ e $\hat{\theta} \in \Theta$, $\hat{\theta}$ satisfaz,

$$\left. \frac{\partial S(\theta)}{\partial \theta_r} \right|_{\hat{\theta}} = 0 \quad (r = 1, 2, \dots, p) \quad (1.12)$$

Para efeito de simplificação utilizamos as seguintes notações:

$$f_i(\theta) = f(x_i; \theta), \quad (1.13)$$

$$\mathbf{f}(\theta) = (f_1(\theta), f_2(\theta), \dots, f_n(\theta))', \quad (1.14)$$

$$\mathbf{F}(\theta) = \frac{\partial \mathbf{f}(\theta)}{\partial \theta'} = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right] \quad (1.15)$$

Também, para brevidade, fazemos,

$$\mathbf{F} = \mathbf{F}(\theta^*) \text{ e } \hat{\mathbf{F}} = \mathbf{F}(\hat{\theta}) \quad (1.16)$$

Assim, podemos escrever,

$$\begin{aligned} S(\theta) &= [\mathbf{y} - \mathbf{f}(\theta)]' [\mathbf{y} - \mathbf{f}(\theta)] \\ &= \|\mathbf{y} - \mathbf{f}(\theta)\|^2 \end{aligned} \quad (1.17)$$

A equação (1.12) conduz à

$$\sum_i \{y_i - f_i(\theta)\} \frac{\partial f_i(\theta)}{\partial \theta_r} \bigg|_{\theta=\hat{\theta}} = 0 \quad (r = 1, 2, \dots, p), \quad (1.18)$$

ou

$$\begin{aligned} \mathbf{0} &= \hat{\mathbf{F}}' \{\mathbf{y} - \mathbf{f}(\hat{\theta})\} \\ &= \hat{\mathbf{F}}' \hat{\boldsymbol{\varepsilon}} \end{aligned} \quad (1.19)$$

Devemos observar que se $\hat{\mathbf{P}}_F = \hat{\mathbf{F}}(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \hat{\mathbf{F}}'$, uma matriz idempotente projetando R^n ortogonalmente em $R[\hat{\mathbf{F}}]$, (ver por exemplo, Seber, 1977), então (1.19) pode ser escrita na forma,

$$\hat{\mathbf{P}}_F \hat{\boldsymbol{\varepsilon}} = \mathbf{0} \quad (1.20)$$

As equações (1.19) são chamadas *equações normais* para o modelo não-linear. Para a maioria dos modelos não-lineares elas não podem ser resolvidas analiticamente, e assim, torna-se necessária a utilização de métodos iterativos para suas resoluções (ver seção (1.4.3)).

EXEMPLO 1.4.1. Considerando o seguinte modelo não-linear,

$$y_i = \alpha x_i^\beta + \varepsilon_i \quad (i = 1, 2, \dots, n), \quad (1.21)$$

as equações normais são dadas por,

$$\sum_i (y_i - \alpha x_i^\beta) x_i^\beta = 0 \quad (1.22)$$

e

$$\sum_i (y_i - \alpha x_i^\beta) \alpha x_i^\beta \log x_i = 0$$

Essas equações não admitem soluções analíticas para α e β e a utilização de um método iterativo, para a resolução dessas equações (1.22), torna-se indispensável.

1.4.2. Aproximação Linear

Consideramos numa vizinhança de θ^* , o verdadeiro valor de θ , a expansão linear de Taylor,

$$f_i(\theta) \approx f_i(\theta^*) + \sum_{r=1}^p \left. \frac{\partial f_i}{\partial \theta_r} \right|_{\theta^*} (\theta_r - \theta_r^*), \quad (1.23)$$

ou

$$f(\theta) \approx f(\theta^*) + F \cdot (\theta - \theta^*), \quad (1.24)$$

onde, de (1.16), $F = F(\theta^*)$; conseqüentemente temos,

$$\begin{aligned} S(\theta) &= \|y - f(\theta)\|^2 \\ &\approx \|y - f(\theta^*) - F \cdot (\theta - \theta^*)\|^2 \\ &= \|z - F \cdot \beta\|^2 \end{aligned} \quad (1.25)$$

onde $z = y - f(\theta^*) = \varepsilon$ e $\beta = \theta - \theta^*$. Das propriedades de modelos lineares (ver por exemplo, Seber, 1977), (1.25) é minimizada quando β é dado por,

$$\hat{\beta} = (F'F)^{-1} F'z$$

Quando n é grande, sob certas condições de regularidade, $\hat{\theta}$ é bem próximo do valor exato na vizinhança de θ^* , desde que $\hat{\theta} - \theta^* \approx \hat{\beta}$ e

$$\hat{\theta} - \theta^* \approx (F'F)^{-1}F'\epsilon \quad (1.26)$$

Além disso, de (1.24) com $\theta = \hat{\theta}$,

$$\begin{aligned} f(\hat{\theta}) - f(\theta^*) &\approx F(\hat{\theta} - \theta^*) \\ &\approx F(F'F)^{-1}F'\epsilon \\ &= P_F \epsilon \end{aligned} \quad (1.27)$$

e

$$\begin{aligned} y - f(\hat{\theta}) &\approx y - f(\theta^*) - F(\hat{\theta} - \theta^*) \\ &\approx \epsilon - P_F \epsilon \\ &= (I_n - P_F) \epsilon \end{aligned} \quad (1.28)$$

onde $P_F = F(F'F)^{-1}F'$ e $I_n - P_F$ são matrizes simétricas e idempotentes (ver por exemplo, Seber, 1977). Desta forma, de (1.28) e (1.27), temos,

$$\begin{aligned} (n-p)s^2 &= S(\hat{\theta}) \\ &= \|y - f(\hat{\theta})\|^2 \\ &\approx \|(I_n - P_F)\epsilon\|^2 \\ &= \epsilon'(I_n - P_F)\epsilon \end{aligned} \quad (1.29)$$

e

$$\begin{aligned} \|f(\hat{\theta}) - f(\theta^*)\|^2 &\approx \|F(\hat{\theta} - \theta^*)\|^2 \\ &= (\hat{\theta} - \theta^*)' F' F (\hat{\theta} - \theta^*) \\ &\approx \|P_F \epsilon\|^2 \\ &= \epsilon' P_F \epsilon \end{aligned} \quad (1.30)$$

Consequentemente, usando (1.29) e (1.30), obtemos,

$$\begin{aligned} S(\theta^*) - S(\hat{\theta}) &\approx \epsilon'\epsilon - \epsilon'(I_n - P_F)\epsilon \\ &= \epsilon' P_F \epsilon \\ &\approx (\hat{\theta} - \theta^*)' F' F (\hat{\theta} - \theta^*) \end{aligned} \quad (1.31)$$

Agora, temos o seguinte teorema:

Teorema (1). Dado $\varepsilon \sim N(0, \sigma^2 I_n)$ e condições de regularidade apropriadas, (Apêndice A), temos para n grande, os seguintes resultados aproximados:

$$i) \hat{\theta} - \theta^* \sim N_p(0, \sigma^2 C^{-1}), \text{ onde } C = F'F = F'(\theta^*)F(\theta^*); \quad (1.32)$$

$$ii) (n-p)s^2 / \sigma^2 \approx \varepsilon'(I_n - P_F)\varepsilon / \sigma^2 \sim \chi_{n-p}^2; \quad (1.33)$$

$$iii) \hat{\theta} \text{ é estatisticamente independente de } s^2; \text{ e} \quad (1.34)$$

$$iv) \frac{[S(\theta^*) - S(\hat{\theta})] / p}{S(\hat{\theta}) / (n-p)} \approx \frac{\varepsilon'P_F\varepsilon}{\varepsilon'(I_n - P_F)\varepsilon} \cdot \frac{n-p}{p} \sim F_{p, n-p} \quad (1.35)$$

A prova do teorema está demonstrada em Seber e Wild, (1989).

Usando *iv*) e (1.31), temos, aproximadamente,

$$\frac{(\hat{\theta} - \theta^*)'F'F(\hat{\theta} - \theta^*)}{ps^2} \sim F_{p, n-p}, \quad (1.36)$$

tal que estimando F por $\hat{F}$, uma região de confiança aproximada $100(1-\alpha)\%$ para θ^* é dada por:

$$\{\theta^*: (\theta^* - \hat{\theta})'\hat{F}'\hat{F}(\theta^* - \hat{\theta}) \leq ps^2 F_{p, n-p}^\alpha\} \quad (1.37)$$

Testes de Hipóteses para θ^* ou subconjuntos de θ^* são desenvolvidos em Seber e Wild, 1989.

1.4.3. Métodos Numéricos

Suponha que $\theta^{(a)}$ é uma aproximação do estimador de mínimos quadrados $\hat{\theta}$ de um modelo não-linear. Vamos usar uma expansão linear de *Taylor*,

$$f(\theta) \approx f(\theta^{(a)}) + F^{(a)}(\theta - \theta^{(a)}) \quad (1.38)$$

onde $F^{(a)} = F(\theta^{(a)})$. Aplicando essa aproximação ao vetor residual $r(\theta)$, obtemos,

$$\begin{aligned} \mathbf{r}(\boldsymbol{\theta}) &= \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}) \\ &\approx \mathbf{r}(\boldsymbol{\theta}^{(a)}) - \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) \end{aligned}$$

Substituindo em $S(\boldsymbol{\theta}) = \mathbf{r}'(\boldsymbol{\theta})\mathbf{r}(\boldsymbol{\theta})$, temos,

$$S(\boldsymbol{\theta}) \approx \mathbf{r}'(\boldsymbol{\theta}^{(a)})\mathbf{r}(\boldsymbol{\theta}^{(a)}) - 2\mathbf{r}'(\boldsymbol{\theta}^{(a)})\mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})' \mathbf{F}^{(a)'} \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}). \quad (1.39)$$

O lado direito da equação acima é minimizado com respeito a $\boldsymbol{\theta}$ quando,

$$\begin{aligned} \boldsymbol{\theta} - \boldsymbol{\theta}^{(a)} &= (\mathbf{F}^{(a)'} \mathbf{F}^{(a)})^{-1} \mathbf{F}^{(a)'} \mathbf{r}(\boldsymbol{\theta}^{(a)}) \\ &= \boldsymbol{\delta}^{(a)} \end{aligned} \quad (1.40)$$

Assim, dado a aproximação $\boldsymbol{\theta}^{(a)}$, a próxima aproximação deve ser dada por,

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + \boldsymbol{\delta}^{(a)} \quad (1.41)$$

Isto produz um método iterativo para a obtenção de $\hat{\boldsymbol{\theta}}$. A aproximação de $S(\boldsymbol{\theta})$ pela equação quadrática (1.39), e os resultados obtidos pelas equações (1.40) e (1.41), referem-se usualmente ao método de Gauss-Newton, que forma a idéia básica de um grande número de algoritmos de mínimos quadrados que são encontrados na literatura. O algoritmo de Gauss-Newton é convergente, isto é $\boldsymbol{\theta}^{(a)} \rightarrow \hat{\boldsymbol{\theta}}$ quando $a \rightarrow \infty$, desde que $\boldsymbol{\theta}^{(1)}$ seja próximo de $\boldsymbol{\theta}^*$ e n seja suficientemente grande. Quando o modelo é linear, isto é, quando $\mathbf{f}(\boldsymbol{\theta})$ é linear nos parâmetros, o algoritmo de Gauss-Newton converge para o ponto de mínimo numa única iteração, partindo-se de qualquer valor inicial. Entretanto, à medida que o grau de não-linearidade do modelo aumenta, então o algoritmo de Gauss-Newton pode não convergir e a função objetivo pode ter mais de um ponto de mínimo e a probabilidade de convergência para o mínimo certo vai diminuindo.

Uma outra possibilidade mais geral, é dada pelo método de Newton, que desenvolve $S(\boldsymbol{\theta})$ diretamente por uma expansão quadrática em série de Taylor. Vamos utilizar as notações,

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (1.42)$$

e

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \quad (1.43)$$

que representam, respectivamente, o vetor gradiente e a matriz Hessiana de $S(\theta)$. Assim, temos a seguinte aproximação quadrática,

$$\begin{aligned} S(\theta) &\approx q_s^{(a)}(\theta) \\ &= S(\theta^{(a)}) + \mathbf{g}'(\theta^{(a)})(\theta - \theta^{(a)}) + \frac{1}{2}(\theta - \theta^{(a)})' \mathbf{H}(\theta^{(a)})(\theta - \theta^{(a)}) \end{aligned} \quad (1.44)$$

que difere de (1.39) somente em $\mathbf{H}(\theta^{(a)})$ que agora é aproximada por $2\mathbf{F}'(\theta^{(a)})\mathbf{F}(\theta^{(a)})$. De qualquer modo, como

$$\frac{\partial^2 S(\theta)}{\partial \theta_r \partial \theta_s} = 2 \sum_{i=1}^n \left\{ \frac{\partial f_i(\theta)}{\partial \theta_r} \frac{\partial f_i(\theta)}{\partial \theta_s} - [y_i - f_i(\theta)] \frac{\partial^2 f_i(\theta)}{\partial \theta_r \partial \theta_s} \right\}, \quad (1.45)$$

então,

$$\mathcal{E} \left[\frac{\partial^2 S(\theta)}{\partial \theta \partial \theta'} \right] = 2\mathbf{F}'(\theta)\mathbf{F}(\theta) \quad (1.46)$$

e $\mathbf{H}(\theta^{(a)})$ é aproximada pelo seu valor esperado calculado em $\theta^{(a)}$ pela equação (1.39).

O mínimo de (1.44) com respeito à θ ocorre quando,

$$\begin{aligned} \theta - \theta^{(a)} &= -[\mathbf{H}(\theta^{(a)})]^{-1} \mathbf{g}(\theta^{(a)}) \\ &= -[\mathbf{H}^{-1} \mathbf{g}]_{\theta=\theta^{(a)}} \end{aligned} \quad (1.47)$$

Este é o método de Newton, e a "correção" $\delta^{(a)}$ em (1.41) agora é dada por (1.47) e é denominada o passo de Newton. Várias modificações para este método têm sido desenvolvidas, baseadas, por exemplo, no escalonamento de $\mathbf{g}$, modificando ou aproximando $\mathbf{H}$ e estimando as primeiras e segundas derivadas em $\mathbf{g}$ e $\mathbf{H}$, usando técnicas de diferenças finitas, (ver por exemplo, Seber e Wild, 1989, para maiores detalhes). Segundo Gill, Murray e Wright (1981) as propriedades de convergência do método de Newton, fazem dele um algoritmo extremamente atraente, sendo freqüentemente tomado como padrão de comparação. Entretanto, o método de Newton (tanto na sua forma mais simples como em qualquer uma de suas modificações), requer o conhecimento das derivadas segundas de $S(\theta)$, o que segundo Jennrich (1979), constitui a sua maior inconveniência ou dificuldade. O método de Newton é extremamente eficiente quando se incorpora uma modificação adequada, mas com a dificuldade em calcular as derivadas ele torna-se impraticável.

Como vimos, a estimação dos parâmetros num sistema não-linear requer o desenvolvimento de técnicas iterativas para a resolução das equações normais não-lineares. Existem, na literatura, alguns outros métodos, tais como, o método steepest descent ou "método do gradiente" e o método de Marquardt (ver por exemplo, Draper e Smith, (1981)).

1.4.4. Estimação de Máxima Verossimilhança

Se a distribuição conjunta de ε_i no modelo (1.10) é assumida conhecida, então o estimador de máxima verossimilhança (e.m.v.) de θ é, por definição, o valor de θ que maximiza a função de verossimilhança. Supondo que ε_i sejam independentes e identicamente distribuídos (i.i.d.) com função densidade $\sigma^{-1}g(\varepsilon/\sigma)$, tal que g seja a distribuição do erro para erros padronizados com variância um, então a função de verossimilhança é dada por,

$$p(y/\theta, \sigma^2) = \prod_{i=1}^n \left[\sigma^{-1} g\left(\frac{y_i - f(x_i; \theta)}{\sigma}\right) \right] \quad (1.48)$$

Considerando o caso onde os erros são normais, vamos calcular o estimador de máxima verossimilhança usando métodos de mínimos quadrados. Para o caso onde não temos normalidade para os erros ver, por exemplo, Seber e Wild (1989).

1.4.4.1. Erros Normais

Se os erros ε_i são i.i.d. $N(0, \sigma^2)$, então (1.48) torna-se,

$$p(y/\theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{[y_i - f(x_i; \theta)]^2}{\sigma^2}\right) \quad (1.49)$$

Ignorando as constantes, denotamos o logaritmo da função de verossimilhança acima por $l(\theta, \sigma^2)$ e obtemos,

$$\begin{aligned} l(\theta, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(x_i; \theta)]^2 \\ &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} S(\theta) \end{aligned} \quad (1.50)$$

Dado σ^2 , (1.50) é maximizada com respeito à θ quando $S(\theta)$ é minimizada, isto é, quando $\theta = \hat{\theta}$ (o estimador de mínimos quadrados). Além disso, $\partial l / \partial \sigma^2 = 0$ tem como solução $\sigma^2 = S(\theta)/n$, que fornece o máximo (para θ dado) se a segunda derivada for negativa, sugerindo que $\hat{\theta}$ e $\hat{\sigma}^2 = S(\hat{\theta})/n$ são os estimadores de máxima verossimilhança, a ser verificado diretamente. Assim, desde que $S(\theta) \geq S(\hat{\theta})$, temos,

$$\begin{aligned} l(\hat{\theta}, \hat{\sigma}^2) - l(\theta, \sigma^2) &= -\frac{n}{2} \log(\hat{\sigma}^2) - \frac{n}{2} - l(\theta, \sigma^2) \\ &\geq -\frac{n}{2} \log \frac{\hat{\sigma}^2}{\sigma^2} - \frac{n}{2} + \frac{1}{2} \frac{S(\hat{\theta})}{\sigma^2} \\ &= -\frac{n}{2} \left(\log \frac{\hat{\sigma}^2}{\sigma^2} + 1 - \frac{\hat{\sigma}^2}{\sigma^2} \right) \\ &\geq 0 \end{aligned}$$

pois $\log x \leq x - 1$ para $x \geq 0$. Como $\hat{\theta}$ e $\hat{\sigma}^2$ maximizam $l(\theta, \sigma^2)$, o valor máximo de (1.49) é dado por,

$$p(y/\hat{\theta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} e^{-n/2} \quad (1.51)$$

Jennrich (1969) menciona que o estimador de mínimos quadrados não é somente um e.m.v. mas, sob condições de regularidade, é também assintoticamente eficiente (ver Seção (12.2), Seber e Wild, 1989). A teoria assintótica de máxima verossimilhança usual não é aplicada diretamente e são necessárias algumas modificações, quando por exemplo y_i não forem identicamente distribuídos, tendo médias diferentes. Mas se $\delta = (\theta, v)'$, onde $v = \sigma^2$, então de (1.50) e (1.46), a matriz informação de Fisher é dada por,

$$\begin{aligned} -E \left[\frac{\partial^2 l}{\partial \delta \partial \delta'} \right] &= \begin{bmatrix} \frac{1}{2\sigma^2} E \left[\frac{\partial^2 S}{\partial \theta \partial \theta'} \right] & -E \left[\frac{\partial^2 l}{\partial \theta \partial v} \right] \\ -E \left[\frac{\partial^2 l}{\partial v \partial \theta'} \right] & -E \left[\frac{\partial^2 l}{\partial v^2} \right] \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma^2} F'(\theta^*) F(\theta^*) & 0 \\ 0' & \frac{n}{2\sigma^4} \end{bmatrix}. \end{aligned} \quad (1.52)$$

Assim, sob condições de regularidade (ver Apêndice A), temos,

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathcal{D}[\sqrt{n}\hat{\theta}] &= \sigma^2 \Omega^{-1} \\ &= \sigma^2 \lim_{n \rightarrow \infty} n [\mathbf{F}'(\theta^*) \mathbf{F}(\theta^*)]^{-1}.\end{aligned}$$

No caso em que temos erros aleatórios i.i.d., a matriz de variância-covariância do e.m.v. $\hat{\theta}$ é dada assintoticamente pelo inverso da matriz de informação de Fisher (1.52).

Seja $\hat{\delta} = (\hat{\theta}', \hat{\sigma}^2)'$, então como $\partial S(\theta)/\partial \theta = 0$ em $\theta = \hat{\theta}$, temos,

$$\left\{ -\frac{\partial^2 l}{\partial \delta \partial \delta'} \right\}_{\hat{\delta}}^{-1} = \begin{bmatrix} \frac{1}{2\hat{\sigma}^2} \left\{ \frac{\partial^2 S}{\partial \delta \partial \delta'} \right\}_{\hat{\theta}} & \mathbf{0} \\ \mathbf{0}' & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}^{-1} = \begin{bmatrix} 2\hat{\sigma}^2 \left\{ \frac{\partial^2 S}{\partial \theta \partial \theta'} \right\}_{\hat{\theta}}^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \quad (1.53)$$

1.5. Contornos de Confiança

Uma idéia da não-linearidade de um modelo pode ser obtida, após a estimação de θ , pelo cálculo da região de confiança elipsoidal obtida, assumindo que a forma linearizada do modelo é válida na vizinhança de $\hat{\theta}$, o estimador final de θ . Uma região de confiança é dada por,

$$(\theta - \hat{\theta})' \hat{\mathbf{F}}' \cdot \hat{\mathbf{F}} \cdot (\theta - \hat{\theta}) \leq p s^2 F_{(p, n-p, 1-\alpha)} \quad (1.54)$$

onde $\mathbf{F} = [(\partial f(x_i; \theta) / \partial \theta_j)]_{\theta^*}$ e $s^2 = S(\hat{\theta}) / (n - p)$. O elipsóide acima não é uma região de confiança verdadeira quando o modelo é não-linear e sim aproximada.

Um contorno de confiança exato é definido considerando $S(\theta) = \text{constante}$, mas como geralmente não conhecemos as propriedades distribucionais no caso não-linear, não conseguimos obter um nível de probabilidade específico. Apesar disso, podemos considerar, por exemplo, o contorno tal que,

$$S(\theta) = S(\hat{\theta}) \left\{ 1 + \frac{p}{n-p} F_{(p, n-p, 1-\alpha)} \right\} \quad (1.55)$$

onde, se o modelo é linear, determina uma região $100(1-\alpha)\%$ exata, e se o modelo é não-linear determina contornos de confiança $100(1-\alpha)\%$ aproximados, que geralmente não são contornos elípticos, apresentam forma irregular e podem apresentar vários mínimos locais. Podemos também construir contornos de confiança baseados nos contornos da função de verossimilhança, via critério da razão de verossimilhanças (ver, por exemplo, Lawless, 1982).

1.6. Comparação de Resultados Assintóticos para Modelos Não-Lineares com Resultados Exatos para Modelos Lineares

1.6.1. Modelos Lineares

Consideramos, por exemplo, o seguinte modelo linear,

$$f(x_i; \theta) = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p \quad (1.56)$$

na forma matricial temos,

$$Y = X\theta + \varepsilon \quad (1.57)$$

onde: Y é o vetor ($n \times 1$) de observações ou variáveis dependentes;

X é a matriz de variáveis independentes ou regressoras;

θ é o vetor ($p \times 1$) de parâmetros desconhecidos;

ε é o vetor ($n \times 1$) de erros aleatórios.

Da teoria linear, sabemos (ver, por exemplo Seber (1977)), que,

i) O estimador de mínimos quadrados $\hat{\theta}$, que minimiza $S(\theta) = \|Y - X\theta\|^2$, é dado por $\hat{\theta} = (X'X)^{-1} X'Y$;

ii) $SQR = S(\hat{\theta}) = Y'Y - \hat{\theta}'X'Y$;

iii) $SQRegressão = \hat{\theta}'X'Y$;

iv) $S(\theta) - S(\hat{\theta}) = (\theta - \hat{\theta})'X'X(\theta - \hat{\theta})$;

v) Supondo: ε_i i.i.d. $N(0, \sigma^2)$, temos os seguintes resultados,

a) $\hat{\theta} \sim N[\theta; (X'X)^{-1}\sigma^2]$, este estimador é não-viciado, normalmente distribuído e tem variância mínima (independentemente do tamanho da amostra);

$$b) S(\hat{\theta}) \sim \sigma^2 \chi_{n-p}^2;$$

$$c) S(\theta) - S(\hat{\theta}) \sim \sigma^2 \chi_p^2;$$

d) $S(\theta) - S(\hat{\theta})$ e $S(\hat{\theta})$ são independentes e portanto, temos,

$$\frac{[S(\theta) - S(\hat{\theta})] / p}{S(\hat{\theta}) / (n - p)} \sim F_{p, n-p} \quad (1.58)$$

Assim, uma região de confiança $100(1 - \alpha)\%$ para $X\theta$ é dada por,

$$S(\theta) = S(\hat{\theta}) \left\{ 1 + \frac{p}{n - p} F_{(p, n-p, 1-\alpha)} \right\}. \quad (1.59)$$

1.6.2. Modelos Não-Lineares

Consideramos o seguinte modelo não-linear,

$$y_i = f(x_i; \theta) + \varepsilon_i, \quad (1.60)$$

onde a função de regressão f não é linear nos parâmetros.

Quando ε é suposto normalmente distribuído, $\hat{\theta}$ não terá distribuição normal, $s^2 = S(\hat{\theta}) / (n - p)$ não será um estimador não-viciado de σ^2 e geralmente não existe matriz de variância-covariância da forma $(X'X)^{-1}\sigma^2$. Apesar disso, as regiões de confiança ainda podem ser definidas pela fórmula,

$$S(\theta) = S(\hat{\theta}) \left\{ 1 + \frac{p}{n - p} F_{(p, n-p, 1-\alpha)} \right\} \quad (1.61)$$

que determina uma região de confiança $100(1 - \alpha)\%$ no modelo linear com erro normal, e no caso não-linear usualmente chamamos essas regiões de confiança aproximadas $100(1 - \alpha)\%$ para θ . Quando a função de regressão f é linear nos parâmetros, os estimadores de mínimos quadrados desses parâmetros são não-viciados e normalmente distribuídos, enquanto que no caso não-linear essas propriedades só são válidas assintoticamente, onde usamos resultados da distribuição assintótica sugerida pela teoria de mínimos quadrados. É bem conhecido que se as observações são normalmente distribuídas, os estimadores de mínimos quadrados coincidem com os estimadores de máxima verossimilhança, portanto sob certas condições de

regularidade são assintoticamente não viciados, eficientes e normalmente distribuídos, assim temos, (Apêndice A),

- i) $\hat{\theta} \stackrel{a}{\sim} N(\theta; \text{Var}(\hat{\theta}))$;
- ii) $\text{Var}(\hat{\theta}) = [\mathbf{F}'(\theta)\mathbf{F}(\theta)]^{-1}\sigma^2$;
- iii) $(n-p)s^2 / \sigma^2 \approx \boldsymbol{\varepsilon}'(\mathbf{I}_n - \mathbf{P}_F)\boldsymbol{\varepsilon} / \sigma^2 \stackrel{a}{\sim} \chi_{n-p}^2$;
- iv) $S(\theta) - S(\hat{\theta}) \approx (\hat{\theta} - \theta)' \mathbf{F}' \cdot \mathbf{F} \cdot (\hat{\theta} - \theta)$;
- v) $\hat{\theta}$ é estatisticamente independente de s^2 ;
- vi) $\frac{[S(\theta) - S(\hat{\theta})] / p}{S(\hat{\theta}) / (n-p)} \approx \frac{\boldsymbol{\varepsilon}' \mathbf{P}_F \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'(\mathbf{I}_n - \mathbf{P}_F)\boldsymbol{\varepsilon}} \cdot \frac{n-p}{p} \stackrel{a}{\sim} F_{p, n-p}$.

Segundo Jennrich (1969), mesmo quando os resíduos não são normalmente distribuídos, mas $n \rightarrow \infty$, o estimador de mínimos quadrados $\hat{\theta}$ ainda apresenta uma distribuição aproximadamente $N(\theta; [\mathbf{F}'(\theta)\mathbf{F}(\theta)]^{-1}\sigma^2)$. As condições de regularidade e as provas de consistência e normalidade assintótica nesse caso, são apresentadas em seu artigo. No caso de amostras pequenas as propriedades desse estimador são desconhecidas, mas à medida que o tamanho amostral n aumenta, os resultados assintóticos vão se tornando mais aplicáveis.

1.7. Problemas Práticos na Estimação de Modelos Não-Lineares

Nesta seção, tratamos de alguns problemas que podem surgir quando ajustamos modelos de regressão não-linear. Problemas tais como: convergência dos métodos iterativos, validade da inferência assintótica na estimação dos parâmetros.

1.7.1. Convergência dos Métodos Iterativos

A maioria dos métodos de estimação, como por exemplo, estimação de mínimos quadrados e máxima verossimilhança, requer um estimador $\hat{\theta}$ obtido através da minimização ou maximização de uma função apropriada $S(\theta)$ de θ . Por exemplo, na estimação por mínimos quadrados, minimizamos $S(\theta) = \|y - f(\theta)\|^2$. Vamos nos referir somente à minimização, pois maximizar $S(\theta)$ é equivalente a minimizar $-S(\theta)$. A função $S(\theta)$ raramente pode ser minimizada analiticamente, assim um procedimento iterativo é usualmente

necessário para encontrar $\hat{\theta}$. Inicialmente temos um valor inicial $\theta^{(1)}$ e sequencialmente encontramos $\theta^{(2)}, \theta^{(3)}, \dots$ até a convergência, onde obtemos uma aproximação para $\hat{\theta}$.

A maioria dos algoritmos que são utilizados para a estimação de θ são adaptações ou aproximações do método clássico de Newton. Não existe um procedimento iterativo que seja superior a todos os outros para todos os problemas, o que existe são certas classes de algoritmos que são, em geral, mais eficientes para determinados tipos de problemas (ou equivalentemente, existem certas classes de problemas para os quais determinados algoritmos são de modo geral mais eficientes).

Desde que as funções de θ sejam aproximadamente quadráticas, os algoritmos usados para a estimação funcionam bem, caso contrário, podemos ter problemas. Algumas vezes, encontramos um mínimo local em lugar do mínimo global. Nenhum dos métodos existentes pode garantir convergência para um mínimo global, entretanto uma maneira de tentar prevenir possíveis soluções locais é através da escolha de bons valores iniciais já que o ponto para o qual um procedimento converge depende também da escolha do valor inicial. Para ilustrar esse fato, observamos o exemplo da figura (1).

Figura 1. Diferentes pontos iniciais podem resultar em diferentes pontos mínimos.

Na prática não devemos nos contentar, pelo menos de imediato, com o valor obtido através do processo iterativo, mas sim, realizar vários testes partindo de vários valores iniciais diferentes e então comparar os resultados; esse procedimento ajuda a prevenir possíveis soluções locais do problema.

O primeiro passo de um processo iterativo, obviamente, é escolher segundo algum critério, o algoritmo a ser utilizado e também uma estimativa inicial para o processo, que podem ser obtidos através de experiências anteriores. Se por acaso, o algoritmo escolhido não convergir satisfatoriamente, podemos tentar reprocessá-lo usando melhores valores iniciais, ou então utilizar uma modificação desse algoritmo ou ainda a utilização de um outro método diferente.

Podemos usar transformações nos parâmetros que, dependendo do problema, possam facilitar a convergência nos métodos iterativos.

Atualmente uma boa opção é utilizar "pacotes", embora também tenham suas limitações. Os pacotes disponíveis para regressão não-linear são, entre outros, *SAS*, *Statgraf*, *BMDP* e o *Splus*.

1.7.2. Validade da Inferência Assintótica

1.7.2.1. Regiões de Confiança

Consideramos o modelo de regressão não-linear usual,

$$y_i = f(x_i; \theta) + \varepsilon_i \quad (i=1, 2, \dots, n) \quad (1.62)$$

onde os ε_i são i.i.d. $N(0, \sigma^2)$. Quando $n \rightarrow \infty$, o estimador de mínimos quadrados $\hat{\theta}$ é assintoticamente $N_p(\theta^*, \sigma^2(\mathbf{F}'\mathbf{F})^{-1})$, onde θ^* é o verdadeiro valor de θ , (ver o teorema (1), da seção (1.4.2)). A matriz $\mathbf{F} = [(\partial f(x_i; \theta) / \partial \theta_j)]_0$ torna-se a matriz $\mathbf{X}$ da teoria linear (Seber, 1977). Em particular, através da analogia com região de confiança linear (ver por exemplo, Seber, 1977), temos,

$$(\beta - \hat{\beta})' \mathbf{X}' \mathbf{X} (\beta - \hat{\beta}) \leq ps^2 F_{(p, n-p, 1-\alpha)} \quad (1.63)$$

Sabemos de (1.37), que,

$$\{\theta: (\theta - \hat{\theta})' \hat{\mathbf{F}}' \cdot \hat{\mathbf{F}} \cdot (\theta - \hat{\theta}) \leq ps^2 F_{(p, n-p, 1-\alpha)}\} \quad (1.64)$$

é uma região de confiança $100(1-\alpha)\%$ aproximada para θ . Aqui $s^2 = S(\hat{\theta}) / (n-p)$, $\hat{\mathbf{F}} = \mathbf{F}(\hat{\theta})$ e $F_{(p, n-p, 1-\alpha)}$ é o valor crítico da distribuição $F_{(p, n-p)}$.

Uma outra forma para a obtenção de regiões de confiança é dada por,

$$\{\theta: S(\theta) \leq cS(\hat{\theta})\} \quad (1.65)$$

para algum c , $c > 1$. Regiões desse tipo são chamadas de regiões de confiança exatas, pois elas não são baseadas em aproximações. Mas, como os níveis de confiança são geralmente desconhecidos, podemos obter níveis aproximados.

Para n grande, $\hat{\theta}$ estará próximo de θ^* , permitindo F . ser substituída por $\hat{F}$.

$$S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)' \hat{F}' \cdot \hat{F} \cdot (\hat{\theta} - \theta^*). \quad (1.66)$$

De (1.31) em (1.64) obtemos a região de confiança (ver Beale, 1960),

$$\left\{ \theta: S(\theta) \leq S(\hat{\theta}) \left(1 + \frac{p}{n-p} F_{(p, n-p, 1-\alpha)} \right) \right\}, \quad (1.67)$$

que é uma região de confiança baseada na normalidade assintótica do estimador de mínimos quadrados $\hat{\theta}$ (neste caso é também o estimador de máxima verossimilhança).

Recentemente não havia nenhum método facilmente aplicável que quantificasse o comportamento não-linear de um modelo e conjunto de dados. A primeira tentativa relevante para medir a não-linearidade foi feita por Beale (1960), mas, essa medida, segundo Guttman e Meeter (1965), tende a subestimar a verdadeira não-linearidade.

Nas seções seguintes, vamos apresentar essas medidas importantes, as medidas de curvatura de Bates e Watts (1980), baseadas no conceito geométrico de curvatura e a medida do vício de Box (1971), que desenvolveu uma fórmula para estimar o vício dos estimadores de máxima verossimilhança. Antes, porém, vamos introduzir o conceito geométrico de curvatura do espaço de estimação.

1.7.2.2. Efeitos de Curvatura

Inicialmente, introduzimos o conceito geométrico de curvatura do espaço de estimação ou locus de solução, através da representação geométrica dos modelos de regressão não-lineares no espaço amostral, em contraste com a dos modelos lineares.

1.7.2.2.1. Representação Geométrica dos Modelos de Regressão Não-Lineares no Espaço Amostral

Consideramos o modelo,

$$Y = f(x_i; \theta) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

onde $E(\varepsilon_i) = 0$ e $\text{Var}(\varepsilon_i) = \sigma^2 \mathbf{I}_n$.

Inicialmente, vamos considerar o caso onde temos $n=2$ e $p=1$, onde p representa o número de parâmetros do modelo. Esse caso, do ponto de vista experimental, é extremamente simples, mas ilustra muito bem o princípio conceitual envolvido. Antes de representar graficamente um modelo não-linear, vamos entender o que ocorre no caso linear.

i) Caso Linear

Seja

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \quad f(x_i; \theta) = \begin{bmatrix} x_1 \theta \\ x_2 \theta \end{bmatrix} \text{ e } \theta \in \mathbb{R}$$

O espaço amostral nesse caso tem dimensão dois e o espaço de estimação ou "locus da solução" é um subespaço de dimensão um do espaço amostral, e é constituído pelos pontos da forma,

$$f(x_i; \theta) = \begin{bmatrix} x_1 \theta \\ x_2 \theta \end{bmatrix},$$

ou seja, é uma reta no $\mathbb{R}^2$ cuja direção é a do vetor $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Uma característica importante dos modelos lineares é que tomando-se valores de θ igualmente espaçados, isto é, $\Delta\theta = \text{constante}$, onde $\Delta\theta = \theta_i - \theta_{i-1}$, $i = 1, 2, \dots$, então os pontos correspondentes no espaço de estimação, $\mathbf{X}\theta_0, \mathbf{X}\theta_1, \mathbf{X}\theta_2, \dots$, também são igualmente espaçados.

Na figura (2) abaixo, consideramos os eixos 1 e 2 como base do espaço amostral e

tomamos $\mathbf{Y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\theta_0 = 1, \theta_1 = 2, \theta_2 = 3, \dots$

Figura 2. Representação dos dados no espaço amostral.

O locus de solução nesse caso é a reta cuja direção é dada pelo vetor $\mathbf{X} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, e os pontos correspondentes a $\theta_0, \theta_1, \theta_2, \dots$ no locus são,

$$\mathbf{X}\theta_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{X}\theta_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{X}\theta_2 = \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \dots$$

Podemos observar que são equidistantes (na figura os valores de $\theta_0, \theta_1, \theta_2, \dots$ são marcados no lugar de $\mathbf{X}\theta_0, \mathbf{X}\theta_1, \mathbf{X}\theta_2, \dots$).

A função

$$S(\theta) = \|\mathbf{Y} - \mathbf{X}\theta\|^2$$

representa o quadrado da distância do ponto $\mathbf{Y}$ a um ponto genérico $\mathbf{X}\theta$ do espaço de estimação. Assim, minimizar $S(\theta)$ em relação a θ corresponde geometricamente a encontrar um ponto no espaço de estimação cuja distância é a menor possível. Na figura (2), o ponto $\hat{\theta}$ (que corresponde ao ponto $\mathbf{X}\hat{\theta}$ nas coordenadas 1 e 2) é o ponto mais próximo de $\mathbf{Y}$, ou seja, é o estimador de mínimos quadrados de θ .

ii) Caso Não-Linear

Vamos considerar o caso onde $n=2$ e $p=1$; o locus de solução não é mais uma reta como no caso anterior, mas uma curva gerada pelos pontos da forma,

$$f(x_i; \theta) = \begin{bmatrix} f(x_1; \theta) \\ f(x_2; \theta) \end{bmatrix}$$

e os pontos do locus de solução correspondentes a $\Delta\theta = \text{constante}$, não são mais necessariamente igualmente espaçados. Como ilustração, consideramos o modelo,

$$Y = f(x_i; \theta) + \varepsilon_i \quad (1.68)$$

onde,

$$Y = \begin{bmatrix} 2,5 \\ 10 \end{bmatrix}, X = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ e } f(x_i; \theta) = \begin{bmatrix} x_1^\theta \\ x_2^\theta \end{bmatrix}.$$

O locus de solução, nesse caso, consiste dos pontos da seguinte forma,

$$f(x_i; \theta) = \begin{bmatrix} 2^\theta \\ 3^\theta \end{bmatrix},$$

onde $\theta \in \mathbb{R}$. Através da figura (3), podemos observar a representação dos dados no espaço amostral.

Figura 3. Representação dos dados no espaço amostral
(Figura retirada de Ratkowsky (1983)).

Observamos através da figura (3), que o locus de solução não é mais uma reta e sim uma curva em torno de $\hat{\theta}$, o valor $\hat{\theta} = 2.0537$ que em termos das coordenadas 1 e 2

corresponde ao ponto $(2^{2.0537}, 3^{2.0537})$, a estimativa de mínimos quadrados de θ . Os pontos do espaço de estimação não são igualmente espaçados como no caso linear, mas apresentam incrementos, $\Delta\theta$, crescentes.

Como os pontos do locus de solução em torno de $\hat{\theta}$ e o tipo de espaçamento existente entre os pontos de locus diferem entre os modelos lineares e os não-lineares, podemos usá-los como medidas de não-linearidade de um modelo. Portanto, quanto maior a curvatura do locus de solução nas vizinhanças de $\hat{\theta}$, isto é, quanto mais o locus se afasta da reta tangente em $\hat{\theta}$ ou ainda se $f(x_i; \theta)$ não é razoavelmente linear na vizinhança de θ , maior será o que Bates e Watts (1980) definem por *não-linearidade intrínseca* do modelo e quanto maior a desigualdade dos espaçamentos entre os pontos, maior será o que Bates e Watts definem por *não-linearidade causada pela parametrização* do modelo.

O termo não-linearidade devido a parametrização depende da maneira como os parâmetros aparecem no modelo, podendo assim, reduzir o grau de não-linearidade através de reparametrizações, enquanto na não linearidade intrínseca o grau de não-linearidade não é alterado através de reparametrizações do modelo.

Consideramos a seguinte reparametrização do modelo (1.68) para ilustrar esse fato,

$$Y = X^{\log \phi} + \varepsilon, \text{ onde } \phi = e^{\theta} \quad (1.69)$$

A figura (4) representa o espaço amostral desse modelo.

Figura 4. Representação dos dados no espaço amostral para o modelo (1.69).

Na figura (4), observamos que a curvatura do locus de solução é a mesma da figura (3), mas o espaçamento entre os pontos correspondentes a $\Delta\theta = 1$ são praticamente iguais

indicando que o grau de não-linearidade causado pela parametrização foi reduzido consideravelmente e portanto espera-se que o modelo (1.69) apresente um comportamento mais próximo do linear nas suas propriedades estatísticas, do que o modelo (1.68), considerando o mesmo conjunto de dados.

No caso onde $n = 3$ e $p = 2$, se o modelo é linear, então o locus de solução é um plano (ver por exemplo, Draper e Smith (1981)) e consequentemente a não-linearidade intrínseca é zero para qualquer conjunto de dados. Além disso, se tomarmos retas paralelas e eqüidistantes no espaço paramétrico, então teremos também retas paralelas e eqüidistantes no locus de solução e portanto a não-linearidade causada pela parametrização também é nula.

A situação é diferente num modelo não-linear, pois o locus de solução é uma superfície curva em torno de $\hat{\theta}$ e retas paralelas e eqüidistantes no espaço paramétrico geram curvas denominadas *linhas paramétricas* que não são mais paralelas e nem eqüidistantes, como ilustração observamos a figura (5).

Figura 5. Representação de um modelo não-linear no caso $n=3$ e $p=2$.

No caso em que temos $n > 3$ e $p \geq 3$ é impossível a representação gráfica do problema, porém o princípio conceitual permanece válido.

Baseados nestes princípios, Bates e Watts (1980) definem medidas quantitativas de não-linearidade que são definidas a partir da especificação do modelo e do conjunto de dados e constituem uma ferramenta de grande valor para a avaliação da não-linearidade. Box (1971) desenvolveu uma fórmula para estimar o vício dos estimadores de máxima

verossimilhança e Ratkowsky (1983) sugere o uso de simulações para verificar o comportamento não-normal dos estimadores e assim, temos condições suficientes para decidir se a utilização da inferência assintótica é válida.

1.7.2.2.2. Medidas de Curvatura de Bates e Watts

Sabemos que se a curvatura é grande, isto significa que o modelo é altamente não-linear e a aproximação linear não é apropriada, assim, precisamos de uma medida para avaliar a não-linearidade do modelo. Bates e Watts (1980) definem medidas de não-linearidade baseadas no conceito geométrico de curvatura.

Segundo Guttman e Meeter (1965), medidas de não-linearidade são expressões que indicam se o grau de não-linearidade, num problema de estimação não-linear, é pequeno suficientemente para justificar a utilização dos resultados usuais da teoria dos modelos lineares como aproximação para os modelos não-lineares.

Consideramos o modelo de regressão da forma,

$$Y = f(x_t; \theta) + \varepsilon_t, \quad t = 1, 2, \dots, n$$

onde $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ é o vetor de parâmetros a serem estimados e a função esperança é dada por,

$$\eta(\theta) = f(x_t; \theta)$$

O procedimento de mínimos quadrados minimiza a soma de quadrados residual,

$$S(\theta) = \sum_i [Y - \eta(\theta)]^2$$

que pode ser escrita como: $S(\theta) = \|Y - \eta(\theta)\|^2$ onde $\eta(\theta) = (\eta_1(\theta), \eta_2(\theta), \dots, \eta_n(\theta))'$.

Os algoritmos para o cálculo de estimativas de mínimos quadrados, tais como o método de Gauss-Newton, são baseados em uma aproximação linear para a função f do modelo. Na vizinhança de um valor fixo do parâmetro θ_0 , a função esperança é aproximada por,

$$\eta(\theta) \cong \eta(\theta_0) + \sum_{i=1}^p (\theta_i - \theta_{i,0}) v_i$$

onde $\mathbf{v}_i = [\dot{v}_i(x_1), \dot{v}_i(x_2), \dots, \dot{v}_i(x_n)]'$ com $\dot{v}_i(x) = \frac{\partial \eta(\theta)}{\partial \theta_i}$ calculado em θ_0 .

Um resultado fundamental de modelos lineares é que as derivadas segundas ou de ordens maiores do que dois da função esperança são iguais a zero. É lógico então, que uma tentativa para medir a não-linearidade de um modelo é dada pela investigação das derivadas segundas da função esperança (Bates, (1978); Bates e Watts, (1980); Beale (1960)).

Podemos então construir a matriz $(n \times p)$ $\dot{V}$, consistindo das primeiras derivadas parciais da função esperança do modelo em relação aos parâmetros, calculada em θ_0 , onde a i -ésima coluna é $\dot{v}_i$. Cada linha de $\dot{V}$ representa o gradiente de uma coordenada de $\eta(\theta)$:

$$\dot{V}_{(n \times p)} = \begin{bmatrix} \frac{\partial \eta_1(\theta)}{\partial \theta_1} & \frac{\partial \eta_1(\theta)}{\partial \theta_2} & \dots & \frac{\partial \eta_1(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_n(\theta)}{\partial \theta_1} & \frac{\partial \eta_n(\theta)}{\partial \theta_2} & \dots & \frac{\partial \eta_n(\theta)}{\partial \theta_p} \end{bmatrix}$$

Os vetores $\dot{v}_i$ $i=1,2,\dots,p$; são denominados vetores tangentes ou vetores velocidade, pois eles fornecem a taxa de mudança de η com respeito à cada parâmetro.

Similarmente, podemos construir a matriz $\ddot{V}$ $(n \times p \times p)$ consistindo das segundas derivadas da função esperança em relação aos parâmetros. Os elementos dessa matriz são,

$$\ddot{v}_{tij} = \frac{\partial^2 \eta_t(\theta)}{\partial \theta_i \partial \theta_j} \quad ; t = 1, \dots, n, \quad i = 1, \dots, p \quad \text{e} \quad j = 1, \dots, p, \quad \text{calculados em } \theta_0,$$

onde cada face $\ddot{v}_t$, $t = 1, \dots, n$ de $\ddot{V}$ é uma matriz $(p \times p)$ de segundas derivadas ou matriz Hessiana de um elemento de $\eta(\theta)$ com respeito a θ .

Podemos escrever $\ddot{v}_{ij} = \frac{\partial \dot{v}_i}{\partial \theta_j}$, $i, j = 1, \dots, p$, como sendo os vetores chamados aceleração, pois eles fornecem a taxa de mudança da velocidade com respeito aos parâmetros.

Bates e Watts (1980) desenvolvem as medidas de não-linearidade a partir de considerações de geometria diferencial.

Em síntese, expressamos uma reta arbitrária no espaço paramétrico, usando um parâmetro geométrico b e passando por θ_0 , da forma,

$$\theta(b) = \theta_0 + bh \quad (1.70)$$

onde $\mathbf{h} = (h_1, h_2, \dots, h_p)'$ é um vetor qualquer diferente de zero. A reta definida em (1.70), gera uma curva dada por,

$$\boldsymbol{\eta}_h(b) = \boldsymbol{\eta}(\boldsymbol{\theta}_0 + b\mathbf{h}). \quad (1.71)$$

A tangente a curva definida em (1.71) e calculada em $b = 0$ é dada por,

$$\dot{\boldsymbol{\eta}}_h = \dot{\mathbf{V}}\mathbf{h},$$

e o conjunto de todas as combinações lineares formam o plano tangente em $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$. A aceleração da linha $\boldsymbol{\eta}_h$ é dada por,

$$\ddot{\boldsymbol{\eta}}_h = \mathbf{h}'\ddot{\mathbf{V}}\mathbf{h},$$

que representa um vetor $(n \times 1)$ de combinação linear da matriz $(p \times p)$ de segundas derivadas parciais para cada uma das n observações.

O vetor aceleração $\ddot{\boldsymbol{\eta}}_h$ pode ser decomposto em:

- $\ddot{\boldsymbol{\eta}}_h^{\text{IN}}$ que determina a mudança na direção do vetor velocidade $\dot{\boldsymbol{\eta}}_h$ normal ao plano tangente;
- $\ddot{\boldsymbol{\eta}}_h^{\text{PE}}$ que determina a mudança na direção do vetor $\dot{\boldsymbol{\eta}}_h$ paralelo ao plano tangente e a mudança na velocidade do ponto em movimento.

Os componentes de aceleração podem ser convertidos em curvaturas dando a curvatura intrínseca,

$$K_h^{\text{IN}} = \frac{\|\ddot{\boldsymbol{\eta}}_h^{\text{IN}}\|}{\|\dot{\boldsymbol{\eta}}_h\|^2}$$

e a curvatura devido aos efeitos dos parâmetros,

$$K_h^{\text{PE}} = \frac{\|\ddot{\boldsymbol{\eta}}_h^{\text{PE}}\|}{\|\dot{\boldsymbol{\eta}}_h\|^2}$$

Somente a última medida de curvatura depende da parametrização escolhida.

K_h^{IN} e K_h^{PE} podem ser convertidas em medidas de curvatura padronizadas γ_h^{IN} e γ_h^{PE} , respectivamente, de tal modo que sejam invariantes por mudança de escala, através da multiplicação pelo raio padronizado $\rho = \hat{\sigma}\sqrt{p}$ onde $\hat{\sigma}$ é a raiz quadrada da variância residual

estimada, $\hat{\sigma}^2$. Isso porque no caso linear a região de confiança $100(1-\alpha)\%$ para $\mathbf{X}\boldsymbol{\theta}$, contida no locus de solução é um círculo (quando $p = 2$) de centro $\mathbf{P} = \mathbf{X}\hat{\boldsymbol{\theta}}$ e raio dado por,

$$\begin{aligned} r &= \sqrt{S(\hat{\boldsymbol{\theta}}) \frac{p}{n-p} F(p, n-p, 1-\alpha)} \\ &= \hat{\sigma} \sqrt{p} \sqrt{F(p, n-p, 1-\alpha)} \end{aligned}$$

(ver por exemplo, Draper e Smith (1981)), ou seja, $r = \rho \sqrt{F}$.

Se o problema é padronizado, dividido por ρ , então o raio de curvatura é simplesmente,

$$r = \sqrt{F(p, n-p, 1-\alpha)}$$

e portanto a curvatura da região de confiança $100(1-\alpha)\%$ é dada por,

$$\frac{1}{\sqrt{F(p, n-p, 1-\alpha)}}.$$

As curvaturas podem ser calculadas diretamente usando a escala dos dados, onde os elementos de $\dot{\mathbf{V}}$ e $\ddot{\mathbf{V}}$ devem ser divididos por ρ .

Os cálculos da curvatura podem ser simplificados fazendo uma transformação de coordenadas do espaço amostral, tal que os p primeiros vetores coordenados sejam paralelos ao plano tangente e os $(n-p)$ últimos vetores sejam ortogonais ao plano tangente. Isto pode ser feito, multiplicando-se todos os vetores no espaço amostral por uma matriz ortogonal $\mathbf{Q}'$, onde $\mathbf{Q}$ é parte da decomposição $\mathbf{QR}$ (ver apêndice B) de $\dot{\mathbf{V}}$, ou seja,

$$\dot{\mathbf{V}}_{(n \times p)} = \mathbf{Q}_{(n \times n)} \mathbf{R}_{(n \times p)} = \mathbf{Q} \begin{bmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{bmatrix} \begin{matrix} \leftarrow (p \times p) \\ \leftarrow (n-p) \times p \end{matrix}$$

onde $\tilde{\mathbf{R}}$ é uma matriz triangular superior e $\mathbf{0}$ é uma matriz de zeros.

As coordenadas no espaço paramétrico também são transformadas de $\boldsymbol{\theta}$ para $\boldsymbol{\phi}$ através da seguinte transformação,

$$\boldsymbol{\phi}_{(p \times 1)} = \tilde{\mathbf{R}}_{(p \times p)} (\boldsymbol{\theta}_{(p \times 1)} - \hat{\boldsymbol{\theta}}_{(p \times 1)})$$

Definindo $\underset{(p \times p)}{L} = \underset{(p \times p)}{\tilde{R}}^{-1}$, a relação entre as matrizes de segundas derivadas nas coordenadas ϕ e θ é dada por,

$$\ddot{U} = \underset{(p \times p)}{L'} \underset{(p \times p)}{\ddot{V}} \underset{(p \times p)}{L}$$

Assim, a matriz $\ddot{A}_{(n \times p \times p)}$ definindo a aceleração é dada por,

$$\ddot{A} = \underset{(n \times n)}{Q'} \underset{(n \times p \times p)}{\ddot{U}} = \ddot{A}^{PE} | \ddot{A}^{IN}$$

onde $\ddot{A}^{PE}$ é a aceleração dos efeitos dos parâmetros consistindo das p primeiras faces de $\ddot{A}$ e $\ddot{A}^{IN}$ é a aceleração intrínseca, consistindo das $(n - p)$ faces de $\ddot{A}$. Para a aceleração intrínseca, a aceleração é de dimensão $(n - p)p^2$, onde a medida de curvatura intrínseca é dada pelo seu máximo (IN). Para a não-linearidade devido a efeito de parâmetros, a aceleração é de dimensão p^3 , e esses p^3 elementos se reduzem a uma única medida dada pelo seu valor máximo (PE).

A significância estatística de IN e PE pode ser obtida comparando esses valores com $1/(2\sqrt{F})$, onde $F = F(p, n - p; 1 - \alpha)$ é obtida da distribuição F e nível de significância α . O valor $1/(\sqrt{F})$ representa o raio de curvatura da região de confiança $100(1 - \alpha)\%$; então o locus de solução pode ser considerado suficientemente linear sobre uma região de confiança $100(1 - \alpha)\%$ se $IN < 1/(2\sqrt{F})$ e analogamente, se $PE < 1/(2\sqrt{F})$. Nesse caso, o modelo apresenta uma não-linearidade intrínseca baixa e uma não-linearidade causada pela parametrização também baixa e portanto seu comportamento na estimação se aproxima muito do comportamento dos modelos lineares e a teoria assintótica é válida para este modelo. Mas se a componente intrínseca é baixa mas $PE > 1/(2\sqrt{F})$, então podemos procurar reparametrizações no modelo que reduzam a componente devido aos parâmetros. Além disso, se a curvatura intrínseca for alta, segundo Ratkowsky (1983), existe pouco propósito em procurar por reparametrizações para reduzir a segunda componente. Neste caso, podemos utilizar a análise Bayesiana usando os métodos numéricos, como por exemplo, Gibbs Sampling e Monte Carlo, para a resolução das integrais.

1.7.2.2.3. Medida do Vício de Box

A fórmula deduzida por Box (1971) para medir o vício nos estimadores de máxima verossimilhança dos parâmetros (no caso das observações serem normalmente distribuídas esses resultados são válidos também para os estimadores de mínimos quadrados) de um modelo de regressão não-linear univariada é dada por,

$$\text{Vício}(\hat{\theta}) = E(\hat{\theta} - \theta) = -\frac{\sigma^2}{2} \left[\sum_{u=1}^n \mathbf{F}_u \mathbf{F}_u' \right]^{-1} \sum_{t=1}^n \mathbf{F}_t \text{tr} \left[\left(\sum_{u=1}^n \mathbf{F}_u \mathbf{F}_u' \right)^{-1} \mathbf{H}_t \right] \quad (1.72)$$

onde:

$$\mathbf{F}_t = \mathbf{F}_u = \begin{bmatrix} \frac{\partial f(x_t; \theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial f(x_t; \theta)}{\partial \theta_p} \end{bmatrix}$$

e

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial^2 f(x_t; \theta)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 f(x_t; \theta)}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x_t; \theta)}{\partial \theta_p \partial \theta_1} & \cdots & \frac{\partial^2 f(x_t; \theta)}{\partial \theta_p^2} \end{bmatrix}$$

Nas aplicações θ e σ^2 são desconhecidos e as estimativas $\hat{\theta}$ e $\hat{\sigma}^2 = S(\hat{\theta}) / (n - p)$ são utilizadas ($S(\hat{\theta})$ é a soma de quadrados dos resíduos).

O vício (1.72) é um vetor ($p \times 1$) representando a discrepância entre os estimadores dos parâmetros e os verdadeiros valores dos parâmetros.

Bates e Watts (1980) demonstraram que o vício de Box está relacionado com a sua medida de não-linearidade causada pela parametrização e portanto, podemos reduzir o vício através do uso de reparametrizações no modelo.

Nas aplicações, a importância da utilização da fórmula de Box é que ela indica qual ou quais parâmetros são os maiores responsáveis por um valor alto da não-linearidade, causada pela parametrização, de Bates e Watts.

Usualmente expressamos o vício dos estimadores em forma de porcentagem,

$$\% \text{ Vício } (\hat{\theta}) = \frac{\text{vício} \times 100}{\hat{\theta}}$$

e como uma lei empírica, devemos considerar os vícios, em valor absoluto, acima de 1% como indicativo de comportamento não-linear.

Box (1971) também desenvolveu uma fórmula para avaliar o vício dos estimadores de uma nova parametrização em função dos vícios dos estimadores da parametrização original. Essa fórmula é dada por,

$$\text{Vício } (\hat{\phi}) = \mathbf{G}' \text{vício}(\hat{\theta}) + \frac{1}{2} \text{tr}[\mathbf{M} \text{cov}(\hat{\theta})] \quad (1.73)$$

onde:

$$\phi = g(\theta),$$

$$\mathbf{G} = \left. \frac{\partial g(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}}, \quad \mathbf{M} = \left. \frac{\partial^2 g(\theta)}{\partial \theta \partial \theta'} \right|_{\theta=\hat{\theta}}$$

$$\text{cov}(\hat{\theta}) = \sigma^2 [\mathbf{F}'(\hat{\theta}).\mathbf{F}(\hat{\theta})]^{-1},$$

e

$$\mathbf{F}(\hat{\theta}) = \left. \frac{\partial f(x_i; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}}.$$

Através da expressão (1.73) podemos avaliar se a reparametrização tem ou não valor, sem a necessidade de refazer todos os cálculos.

A variância de $\hat{\phi}$ pode ser obtida através da matriz de covariância assintótica $\sigma^2 [\mathbf{F}'(\hat{\theta}).\mathbf{F}(\hat{\theta})]^{-1}$, da seguinte maneira,

$$\text{Var } (\hat{\phi}) = \text{tr}[(\mathbf{G}\mathbf{G}') \text{cov}(\hat{\theta})].$$

Devemos ressaltar que além da avaliação da não-linearidade de um modelo, através das medidas de curvatura de Bates e Watts, podemos realizar estudos com base em simulação, o que segundo *Ratkowsky (1983)* é uma das melhores maneiras de avaliar as propriedades dos estimadores, (ver Apêndice C).

Em apêndice, temos um programa em linguagem *Fortran* (fornecido por *Ratkowsky, 1983*), que calcula as medidas de curvatura de Bates e Watts, o vício de Box e calcula as estimativas de mínimos quadrados dos parâmetros de modelos não-lineares, utilizando o método de Gauss-Newton.

Capítulo 2

INFERÊNCIA CLÁSSICA

Um problema importante da inferência estatística é a estimação dos parâmetros de um modelo deduzidos de um conjunto de dados amostrais. No capítulo (1), apresentamos uma abordagem da estimação pontual dos parâmetros de um modelo não-linear. Na prática, também temos interesse em tomar decisões sobre os parâmetros baseados nas informações das amostras; assim, torna-se extremamente necessário a formulação de hipóteses estatísticas sobre os parâmetros.

Neste capítulo, vamos discutir, sob o enfoque clássico, intervalos de confiança e testes de hipóteses para os modelos não-lineares. Devemos novamente ressaltar que a validade da inferência assintótica depende muito do grau de não-linearidade do modelo e de uma boa parametrização.

2.1. Intervalos de Confiança Assintóticos

Consideramos o modelo não-linear,

$$y_i = f(x_i; \theta) + \varepsilon_i \quad (i = 1, 2, \dots, n), \quad (2.1)$$

onde ε_i são i.i.d. $N(0, \sigma^2)$. Por conveniência de notação vamos substituir θ^* , o verdadeiro vetor de parâmetros, pelo vetor de parâmetros θ . Se estamos interessados em construir um intervalo de confiança para uma dada combinação linear $a'\theta$, então podemos aplicar o teorema (1) da seção (1.4.2), assim, temos o seguinte resultado assintótico,

$$\hat{\theta} \sim N_p(\theta, \sigma^2 C^{-1}), \quad C = F'F, \quad (2.2)$$

$$\text{onde } F = F(\theta) = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right], \quad j = 1, 2, \dots, p.$$

Do resultado (2.2) acima, temos assintoticamente, que $\mathbf{a}'\hat{\boldsymbol{\theta}} \sim N(\mathbf{a}'\boldsymbol{\theta}, \sigma^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a})$ independentemente de $s^2 = \|\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}})\|^2 / (n - p)$, onde s^2 é um estimador não-viciado de σ^2 e p representa o número de parâmetros. Para n grande, temos aproximadamente,

$$T = \frac{\mathbf{a}'\hat{\boldsymbol{\theta}} - \mathbf{a}'\boldsymbol{\theta}}{s(\mathbf{a}'\mathbf{C}^{-1}\mathbf{a})^{1/2}} \sim t_{n-p}, \quad (2.3)$$

onde t_{n-p} é a distribuição t de Student com $(n - p)$ graus de liberdade. Podemos construir um intervalo de confiança $100(1-\alpha)\%$ aproximado para $\mathbf{a}'\boldsymbol{\theta}$ da seguinte forma,

$$\mathbf{a}'\hat{\boldsymbol{\theta}} \pm t_{n-p}^{\alpha/2} s(\mathbf{a}'\mathbf{C}^{-1}\mathbf{a})^{1/2}. \quad (2.4)$$

Neste caso, $\mathbf{C}$ pode ser estimada por $\hat{\mathbf{C}} = \hat{\mathbf{F}}'\hat{\mathbf{F}}$. (Donaldson e Schnabel (1987)). Fazendo $\mathbf{a}' = (0, 0, \dots, 0, 1, 0, \dots, 0)$, onde o r -ésimo elemento de $\mathbf{a}$ é igual a um e os demais elementos são iguais a zero, e definindo $[(\hat{c}^{rr})] = \hat{\mathbf{C}}^{-1}$, um intervalo de confiança para o r -ésimo elemento de $\boldsymbol{\theta}$, θ_r , é dado por,

$$\hat{\theta}_r \pm t_{n-p}^{\alpha/2} s \sqrt{\hat{c}^{rr}}. \quad (2.5)$$

2.2. Intervalos de Confiança para Funções Não-Lineares de $\boldsymbol{\theta}$

Intervalos de confiança para funções não-lineares de $\boldsymbol{\theta}$, $a(\boldsymbol{\theta})$, podem ser construídos de forma similar, utilizando a expansão linear em série de *Taylor*:

$$a(\hat{\boldsymbol{\theta}}) \approx a(\boldsymbol{\theta}) + \mathbf{a}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \quad (2.6)$$

onde $\mathbf{a}' = (\partial a / \partial \theta_1, \partial a / \partial \theta_2, \dots, \partial a / \partial \theta_p)$. Assim,

$$\begin{aligned} \mathcal{Q}[a(\hat{\boldsymbol{\theta}})] &\approx \mathcal{Q}[\mathbf{a}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})] \\ &= \mathbf{a}'\mathcal{Q}[\hat{\boldsymbol{\theta}}]\mathbf{a} \\ &\approx \sigma^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a} \end{aligned}$$

e de (2.2), temos o resultado assintótico,

$$a(\hat{\theta}) \sim N(a(\theta), \sigma^2 a'(\theta) C^{-1} a).$$

Portanto, um intervalo de confiança $100(1-\alpha)\%$ aproximado para $a(\theta)$ é dado por,

$$a(\hat{\theta}) \pm t_{n-p}^{\alpha/2} s(a'(\hat{\theta}) C^{-1} a(\hat{\theta}))^{1/2} \quad (2.7)$$

2.3. Intervalos de Confiança para Previsão

Utilizando a linearização assintótica de (2.1), podemos aplicar alguns métodos lineares existentes, para encontrar um intervalo de previsão para y em $x = x_0$. Seja,

$$y_0 = f(x_0; \theta) + \varepsilon_0,$$

onde $\varepsilon_0 \sim N(0, \sigma^2)$ e independente de ε , uma estimativa óbvia de y_0 é $\hat{y}_0 = f(x_0; \hat{\theta})$. Como para n grande, $\hat{\theta}$ é próximo do verdadeiro valor de θ , temos a expansão usual de *Taylor*,

$$f(x_0; \hat{\theta}) \approx f(x_0; \theta) + f'_0(\hat{\theta} - \theta),$$

onde

$$f'_0 = \left(\frac{\partial f(x_0; \theta)}{\partial \theta_1}, \frac{\partial f(x_0; \theta)}{\partial \theta_2}, \dots, \frac{\partial f(x_0; \theta)}{\partial \theta_p} \right).$$

Assim,

$$y_0 - \hat{y}_0 \approx y_0 - f(x_0; \theta) - f'_0(\hat{\theta} - \theta) = \varepsilon_0 - f'_0(\hat{\theta} - \theta).$$

De (2.2) e da independência de $\hat{\theta}$ e ε_0 ,

$$\begin{aligned} E[y_0 - \hat{y}_0] &\approx E[\varepsilon_0] - f'_0 E[\hat{\theta} - \theta] \approx 0, \\ \text{Var}[y_0 - \hat{y}_0] &\approx \text{Var}[\varepsilon_0] + \text{Var}[f'_0(\hat{\theta} - \theta)] \\ &\approx \sigma^2 + \sigma^2 f'_0 (F' \cdot F)^{-1} f_0 \\ &= \sigma^2 (1 + \nu_0), \end{aligned}$$

e $y_0 - \hat{y}_0$ é assintoticamente $N(0, \sigma^2[1 + \nu_0])$; s^2 é independente de y_0 e é assintoticamente independente de $\hat{\theta}$, isto é, s^2 é assintoticamente independente de $y_0 - \hat{y}_0$. Assim, temos assintoticamente,

$$\frac{y_0 - \hat{y}_0}{s\sqrt{1 + v_0}} \sim t_{n-p},$$

e um intervalo de confiança $100(1-\alpha)\%$ aproximado para y_0 é dado por,

$$\hat{y}_0 \pm t_{n-p}^{\alpha/2} s [1 + \mathbf{f}_0'(\mathbf{F}'\mathbf{F})^{-1}\mathbf{f}_0]^{1/2}. \quad (2.8)$$

Desde que $\mathbf{f}_0$ e $\mathbf{F}$ são funções de $\boldsymbol{\theta}$, v_0 é uma função de $\boldsymbol{\theta}$ e podemos estimá-la substituindo $\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}$.

2.4. Regiões de Confiança

Na seção (1.7), encontramos uma breve discussão sobre regiões de confiança. Considerando $S(\boldsymbol{\theta}) = \|\boldsymbol{\varepsilon}(\boldsymbol{\theta})\|^2 = \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|^2$, uma região de confiança é dada por,

$$\left\{ \boldsymbol{\theta} : \frac{S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}})}{S(\hat{\boldsymbol{\theta}})} \leq \frac{p}{n-p} F_{p, n-p}^{\alpha} \right\}, \quad (2.9)$$

ou, se escrevemos $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}(\hat{\boldsymbol{\theta}})$, temos então a região de confiança,

$$\{ \boldsymbol{\theta} : \|\boldsymbol{\varepsilon}(\boldsymbol{\theta})\|^2 \leq (1+f) \|\hat{\boldsymbol{\varepsilon}}\|^2 \}, \quad f = \frac{p}{n-p} F_{p, n-p}^{\alpha}. \quad (2.10)$$

Uma versão aproximada de (2.9) é proposta por Box e Coutie (1956). Expandindo $S(\boldsymbol{\theta})$ como em (1.44), temos,

$$\begin{aligned} S(\boldsymbol{\theta}) &\approx S(\hat{\boldsymbol{\theta}}) + \left(\frac{\partial S}{\partial \boldsymbol{\theta}'} \right)_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \left(\frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \\ &= S(\hat{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \hat{\mathbf{H}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \end{aligned} \quad (2.11)$$

desde que $\partial S / \partial \boldsymbol{\theta} = \mathbf{0}$ em $\hat{\boldsymbol{\theta}}$ e $\hat{\mathbf{H}} = \left(\frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)$, calculada em $\hat{\boldsymbol{\theta}}$. Então (2.9) torna-se, aproximadamente,

$$\theta(\theta - \hat{\theta})' \hat{H}(\theta - \hat{\theta}) \leq 2s^2 p F_{p, n-p}^{\alpha}, \quad (2.12)$$

onde $s^2 = S(\hat{\theta}) / (n - p)$. Devemos ressaltar que a validade da teoria assintótica acima depende muito do grau de não-linearidade do modelo e da escolha dos parâmetros θ . Se a aproximação linear usada é satisfatória, então os resultados da teoria linear podem ser aplicados com $\mathbf{X} = \hat{\mathbf{F}}$. Os métodos lineares para regiões de confiança e intervalos de confiança simultâneos são discutidos por Seber (1977), Miller (1981) e Savin (1984). Se os efeitos de curvatura não podem ser ignorados, então algumas modificações na teoria acima devem ser feitas, (ver por exemplo, Seber e Wild, 1989). Quando temos interesse nos parâmetros originais e a curvatura devido ao efeito dos parâmetros for alta, podemos por exemplo, utilizar o teste da razão de verossimilhanças ou o teste do "escore" que são invariantes a reparametrizações.

2.5. Métodos Assintóticos baseados na Função de Verossimilhança

Supor que $l(\gamma)$ seja o logaritmo da função de verossimilhança para um modelo geral com vetor paramétrico γ r -dimensional. Então, sob certas condições usuais de regularidade, podemos testar a hipótese $H_0 : \gamma = \gamma_0$ usando a seguinte estatística,

$$\begin{aligned} \text{LR} &= -2 \log(\lambda) \\ &= 2[l(\hat{\gamma}) - l(\gamma_0)] \end{aligned} \quad (2.13)$$

onde λ é a estatística da razão de verossimilhanças e $\hat{\gamma}$ é o e.m.v. de γ , maximizando $l(\gamma)$. A estatística LR é assintoticamente χ_r^2 quando a hipótese H_0 é verdadeira. Uma região de confiança $100(1-\alpha)\%$ aproximada para γ é dada por,

$$\{\gamma : 2[l(\hat{\gamma}) - l(\gamma)] \leq \chi_r^2(\alpha)\}, \quad (2.14)$$

onde $P[\chi_r^2 \geq \chi_r^2(\alpha)] = \alpha$. Usando um argumento similar, se $\gamma = (\gamma_1', \gamma_2')'$ e $\tilde{\gamma}_1(\gamma_2)$ é o e.m.v. de γ_1 dado o valor de γ_2 r_2 -dimensional, então temos a região de confiança para γ_2 :

$$\{\gamma_2 : 2[l(\hat{\gamma}) - l(\hat{\gamma}_1(\gamma_2), \gamma_2)] \leq \chi_{r_2}^2(\alpha)\}. \quad (2.15)$$

Para maiores informações ver Seber e Wild (1989).

2.6. Testes de Hipóteses

Os processos que habilitam a decidir se aceitamos ou rejeitamos as hipóteses formuladas sobre os parâmetros do modelo são denominados testes de hipóteses. Para tanto, podemos estudar dois tipos de hipóteses: a hipótese linear e a hipótese não-linear.

2.6.1. Hipóteses Lineares

Supondo que temos interesse em testar $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$. Dispomos de dois métodos assintóticos que são equivalentes. O primeiro método utiliza a fórmula,

$$F_1 = (\hat{\theta} - \theta_0)' \hat{C} (\hat{\theta} - \theta_0) / ps^2, \quad (2.16)$$

onde $\hat{C} = \hat{F}' \cdot \hat{F}$. e F_1 é assintoticamente $F_{p, n-p}$, quando a hipótese H_0 é verdadeira. Rejeitamos H_0 ao nível de significância α se $F_1 > F_{p, n-p}^\alpha$. O segundo método utiliza a estatística da razão de verossimilhanças, que, de (1.51) e (1.50) (fazendo $\tilde{\sigma}_0^2 = S(\theta_0)/n$), é dada por,

$$\begin{aligned} \lambda &= \frac{\sup_{\sigma^2} p(y; \theta_0, \sigma^2)}{\sup_{\sigma^2, \theta} p(y; \theta, \sigma^2)} \\ &= \frac{(2\pi \tilde{\sigma}_0)^{-n/2} e^{-n/2}}{(2\pi \hat{\sigma}_0)^{-n/2} e^{-n/2}} \\ &= (\hat{\sigma}^2 / \tilde{\sigma}_0^2)^{n/2} \\ &= [S(\hat{\theta}) / S(\theta_0)]^{n/2} \end{aligned} \quad (2.17)$$

Rejeitamos a hipótese H_0 se λ for muito pequena, ou equivalentemente, se

$$F_2 = (\lambda^{-2/n} - 1) \frac{n-p}{p} = \frac{S(\theta_0) - S(\hat{\theta})}{S(\hat{\theta})} \frac{n-p}{p} \quad (2.18)$$

for muito grande. De (1.35) vimos que F_2 é assintoticamente $F_{p, n-p}$ quando a hipótese H_0 é verdadeira. Consequentemente, devemos rejeitar H_0 se θ_0 cair fora da região de confiança (2.9). O poder deste teste é discutido por Gallant (1975;a,b).

Supondo agora a partição $\theta = (\theta_1', \theta_2')'$, onde θ_2 é um vetor $(p_2 \times 1)$, e que desejamos testar $H_{02}: \theta_2 = \theta_{02}$. Neste caso, os dois procedimentos poderiam ser utilizados. Particionando $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')'$ e

$$C^{-1} = \begin{pmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \hat{C}^{11} & \hat{C}^{12} \\ \hat{C}^{21} & \hat{C}^{22} \end{pmatrix},$$

observamos que de (2.2) $\hat{\theta}_2 \sim N_{p_2}(\theta_2, \sigma^2 \hat{C}^{22})$, aproximadamente. Particionando $\hat{F} = (\hat{F}_1, \hat{F}_2)$, onde $\hat{F}_2$ é $(n \times p_2)$, temos,

$$\begin{aligned} (\hat{C}^{22})^{-1} &= \hat{C}_{22} - \hat{C}_{21} \hat{C}_{11}^{-1} \hat{C}_{12} \\ &= \hat{F}_2' \hat{F}_2 - \hat{F}_2' \hat{F}_1 (\hat{F}_1' \hat{F}_1)^{-1} \hat{F}_1' \hat{F}_2 \\ &= \hat{F}_2' (I_n - \hat{P}_{F_1}) \hat{F}_2 \end{aligned} \quad (2.19)$$

Analogamente a (2.16), temos a estatística para o teste,

$$\tilde{F}_1 = (\hat{\theta}_2 - \theta_{02})' (\hat{C}^{22})^{-1} (\hat{\theta}_2 - \theta_{02}) / p_2 s^2. \quad (2.20)$$

Essa estatística tem aproximadamente uma distribuição $F_{p_2, n-p}$, quando a hipótese H_0 é verdadeira. A regra de decisão é rejeitar H_0 se $\tilde{F}_1 > F_{p_2, n-p}^\alpha$.

Para aplicar o procedimento da razão de verossimilhanças, precisamos minimizar $S(\theta) [= S(\theta_1, \theta_2)]$ sujeita à restrição $\theta_2 = \theta_{02}$. Seja $S(\tilde{\theta}) = S[\tilde{\theta}_1(\theta_{02}), \theta_{02}]$ o valor mínimo de $S(\theta)$ sob H_{02} , então, a estatística da razão de verossimilhanças é,

$$\lambda = [S(\hat{\theta}) / S(\tilde{\theta})]^{n/2}. \quad (2.21)$$

Usando (2.18), temos o teste estatístico dado na forma,

$$\tilde{F}_2 = \frac{S(\tilde{\theta}) - S(\hat{\theta})}{S(\hat{\theta})} \frac{n-p}{p_2}, \quad (2.22)$$

onde $\tilde{F}_2$ tem aproximadamente uma distribuição $F_{p_2, n-p}$ quando H_0 é verdadeira.

Um terceiro método para a construção de testes de hipóteses é o chamado método dos Multiplicadores de Lagrange que está disponível em duas versões, fornecendo os seguintes testes estatísticos,

$$\tilde{F}_3 = \frac{\{y - f(\tilde{\theta})\}' [F \cdot (F' \cdot F)^{-1} F']_{\tilde{\theta}} \{y - f(\tilde{\theta})\} / p_2}{S(\tilde{\theta}) / (n - p)}$$

e

$$LM = \frac{\{y - f(\tilde{\theta})\}' [F \cdot (F' \cdot F)^{-1} F']_{\tilde{\theta}} \{y - f(\tilde{\theta})\}}{S(\tilde{\theta}) / n} = \frac{\tilde{\varepsilon}' \tilde{P}_F \tilde{\varepsilon}}{\tilde{\varepsilon}' \tilde{\varepsilon} / n}$$

onde $\tilde{\varepsilon} = y - f(\tilde{\theta})$. Neste caso, $\tilde{F}_3$ é assintoticamente equivalente à $\tilde{F}_2$, assim $\tilde{F}_3 \stackrel{a}{\sim} F_{p_2, n-p}$ quando a hipótese H_o é verdadeira. A estatística LM é a forma usual do teste do multiplicador de Lagrange e é assintoticamente $\chi^2_{p_2}$, quando a hipótese H_o é verdadeira. Essa estatística tem uma vantagem: somente é necessário o estimador restrito $\tilde{\theta}$ para o seu cálculo. Usando matrizes de projeções (ver Seber e Wild, 1989) podemos mostrar (Gallant, 1987:p.87) que,

$$R = \frac{n^{-1}LM}{1 - n^{-1}LM} \frac{n-p}{p_2} \stackrel{a}{\sim} F_{p_2, n-p} \quad (2.23)$$

quando a hipótese H_o é verdadeira. Assim, podemos usar o teste LM com um valor crítico dado por,

$$d_\alpha = \frac{np_2 F_\alpha}{n - p + p_2 F_\alpha}, \quad F_\alpha = F_{p_2, n-p}^\alpha \quad (2.24)$$

Apesar de $\tilde{F}_3$ requerer os dois estimadores $\hat{\theta}$ e $\tilde{\theta}$, Gallant (1987) mostra que esta estatística tem teoria distribucional simples e melhor do que a estatística LM.¹

Uma outra possibilidade para a construção de testes para modelos não-lineares, é dada por Hamilton (1986) que considera a estatística do "escore eficiente":

$$\tilde{F}_4 = \frac{\tilde{\varepsilon}' (\tilde{P}_F - \tilde{P}_{F_1}) \tilde{\varepsilon} / p_2}{\tilde{\varepsilon}' (\mathbf{I}_n - \tilde{P}_F) \tilde{\varepsilon} / (n - p)} \quad (2.25)$$

¹Para maiores informações sobre as estatísticas LM, $\tilde{F}_2$ e $\tilde{F}_3$, ver por exemplo, Gallant (1987).

onde $\tilde{\mathbf{P}}_F = [\mathbf{F}_1(\mathbf{F}_1'\mathbf{F}_1)^{-1}\mathbf{F}_1']_{\theta}$ e $\mathbf{F}_1$ consiste das p_1 primeiras ($p_1 = p - p_2$) colunas de $\mathbf{F}$. (que corresponde a $\partial \mathbf{f} / \partial \theta_1$). Quando a hipótese H_0 é verdadeira, temos que $\tilde{F}_4 \rightarrow R$, dado em (2.23) quando $n \rightarrow \infty$. Consequentemente, a distribuição de $\tilde{F}_4$ é aproximadamente $F_{p_2, n-p}$.

A estatística $\tilde{F}_1$ pode ser seriamente afetada por qualquer efeito de curvatura paramétrica e seu uso indiscriminado não é recomendado.²

Testes para hipóteses não-lineares podem ser vistos em Seber e Wild (1989).

2.7. Análise Clássica para o Modelo de Michaelis-Menten baseada na Normalidade Assintótica

O uso da aproximação Normal assintótica dos estimadores de máxima verossimilhança é um dos procedimentos mais difundidos na aplicação de técnicas estatísticas em problemas práticos.

Nesta seção, vamos determinar os estimadores de máxima verossimilhança para os parâmetros do modelo de Michaelis-Menten e a partir da normalidade assintótica dos estimadores de máxima verossimilhança obter as inferências sobre os parâmetros do modelo ou funções dos mesmos. Mas, é importante salientar, que devemos, sempre, verificar se essa aproximação é viável, estudando sua precisão. Para este propósito, vamos utilizar algumas medidas de não-normalidade de funções de verossimilhança recentes, propostas por Hills e Smith (1993) e por Kass e Slate (1992), explorando o uso de algumas reparametrizações, que serão apresentadas no próximo capítulo, para obter melhores aproximações Normais para a função de verossimilhança dos parâmetros envolvidos.

Devemos ressaltar, que uma outra alternativa para a obtenção de inferências precisas sobre os parâmetros é o uso de métodos Bayesianos, que exploraremos em capítulos posteriores.

2.7.1. Formulação do Modelo

Um modelo não-linear muito popular considerado em vários experimentos bioquímicos e farmacológicos é o modelo de Michaelis-Menten, dado por,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i \quad (2.26)$$

²Maiores detalhes sobre as estatísticas ver Seber e Wild (1989).

onde $y_i, i = 1, 2, \dots, n$ é a variável dependente, x_i é a variável independente, θ_1 e θ_2 são parâmetros do modelo e os erros ε_i são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal $N(0, \sigma^2)$.

Uma aplicação especial do modelo (2.26) é dado em estudos de cinética de enzimas, por exemplo, para descrever o metabolismo de substâncias no fígado, onde relaciona a velocidade inicial y_i de uma reação enzimática com a concentração x_i de um substrato. Neste caso, desejamos estimar $\theta_1 = V$ ou $V_{\max}$, a velocidade máxima e $\theta_2 = k_m$ a constante de Michaelis. Neste contexto, o modelo (2.26) é conhecido como *equação de Michaelis-Menten* (ver por exemplo, Raaijmakers, 1987).

2.7.2. A Função de Verossimilhança

A partir do momento em que observamos os dados, toda a informação sobre os parâmetros está contida na função de verossimilhança.

A função de verossimilhança para os parâmetros θ_1, θ_2 e σ^2 do modelo de Michaelis-Menten (2.26), considerando uma amostra aleatória de tamanho n é dada na forma,

$$L(\theta_1, \theta_2, \sigma^2) = \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right\}$$

Isto é,

$$L(\theta_1, \theta_2, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right\} \quad (2.27)$$

O logaritmo da função de verossimilhança (2.27) é dado por,

$$l(\theta_1, \theta_2, \sigma^2) \propto -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2 \quad (2.28)$$

2.7.3. Estimadores de Máxima Verossimilhança

Geralmente, nosso objetivo é obter informações a respeito dos parâmetros, isto é, obter informação sobre características numéricas desconhecidas da população. Um estimador muito utilizado devido às suas boas propriedades é dado pelo estimador de máxima verossimilhança.

Os estimadores de máxima verossimilhança para os parâmetros do modelo (2.26), obtidos das equações de verossimilhança,

$$\frac{\partial \ell(\theta_1, \theta_2, \sigma^2)}{\partial \theta_1} = -\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} + \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} = 0,$$

$$\frac{\partial \ell(\theta_1, \theta_2, \sigma^2)}{\partial \theta_2} = -\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} + \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} = 0$$

e

(2.29)

$$\frac{\partial \ell(\theta_1, \theta_2, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 = 0$$

são dados pelas equações,

$$\hat{\theta}_1 \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^2} - \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)} = 0,$$

$$\hat{\theta}_1 \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^3} - \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^2} = 0$$

(2.30)

e

$$\hat{\sigma}^2 = \frac{S(\hat{\theta}_1, \hat{\theta}_2)}{n}$$

onde $S(\theta_1, \theta_2) = \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2$. Devemos observar, que para a obtenção dos e.m.v. de θ_1 e θ_2 necessitamos do uso de um método iterativo, por exemplo, o método de Newton-Raphson, pois este sistema é não-linear em θ e $\hat{\theta}$ não é obtido explicitamente.

2.7.3.1. Inferências sobre os Parâmetros do modelo

A partir dos estimadores de máxima verossimilhança $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\sigma}^2$ podemos obter intervalos de confiança e testes de hipóteses para os parâmetros do modelo, usando a teoria assintótica para grandes amostras baseada na matriz de informação de Fisher ou informação observada. Assim, consideramos,

$$(\hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) \stackrel{a}{\sim} N\{(\theta_1, \theta_2, \sigma^2); I^{-1}(\hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2)\} \quad (2.31)$$

onde $I(\theta_1, \theta_2, \sigma^2)$ é a matriz de informação de Fisher (ver (1.52)), dada por,

$$I(\theta_1, \theta_2, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} F'(\theta) \cdot F(\theta) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \quad (2.32)$$

onde $F(\theta) = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right) \right]$, $i = 1, 2, \dots, n$, $j = 1, 2$ e $f_i(\theta) = \frac{\theta_1 x_i}{\theta_2 + x_i}$.

Assim,

$$I(\theta_1, \theta_2, \sigma^2) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} & -\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} & 0 \\ -\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} & \theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} & 0 \\ 0 & 0 & \frac{n}{2\sigma^2} \end{bmatrix} \quad (2.33)$$

e $I^{-1}(\theta_1, \theta_2, \sigma^2)$, o inverso da matriz de informação de Fisher, dada em (2.31), representa a matriz de variância e covariância assintótica de $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\sigma}^2$. Quando os elementos da matriz (2.33) são constantes, temos uma boa aproximação normal assintótica para os estimadores de máxima verossimilhança $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\sigma}^2$ (ver por exemplo, Sprott, 1973 e 1980) e as inferências sobre θ_1, θ_2 e σ^2 podem ser baseadas em (2.31).

Podemos, facilmente, construir intervalos de confiança para os parâmetros do modelo (2.26). Por exemplo, para obter um intervalo de confiança para θ e σ , considerando um nível de confiança $(1-\alpha)$, temos,

$$P\left[\hat{\theta}_i - z_{\alpha/2}\sqrt{I_{ii}^{-1}(\theta, \sigma^2)} \leq \theta_i \leq \hat{\theta}_i + z_{\alpha/2}\sqrt{I_{ii}^{-1}(\theta, \sigma^2)}\right] = 1 - \alpha$$

e

$$P\left[\hat{\sigma}^2 - z_{\alpha/2}\sqrt{I_{33}^{-1}(\theta, \sigma^2)} \leq \sigma^2 \leq \hat{\sigma}^2 + z_{\alpha/2}\sqrt{I_{33}^{-1}(\theta, \sigma^2)}\right] = 1 - \alpha$$

onde $I_{ii}^{-1}(\theta_1, \theta_2, \sigma^2)$ representa o i -ésimo elemento diagonal da matriz de variância e covariância assintótica de $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\sigma}^2$ e $z_{\alpha/2}$ é um percentil da distribuição normal padronizada obtido de uma tabela estatística.

A precisão desses resultados aproximados podem ser melhorados usando uma reparametrização apropriada. Dependendo do modelo e conjunto de dados, podemos ter uma alta não-linearidade paramétrica, necessitando assim, de reparametrização. No próximo capítulo estudamos alguns aspectos de reparametrização com o intuito de obter resultados assintóticos clássicos precisos.

Capítulo 3

ALGUNS ASPECTOS DE REPARAMETRIZAÇÃO

Usualmente, as análises estatísticas clássicas para os modelos não-lineares são baseadas em resultados assintóticos que dependem de uma boa parametrização. A eficiência destes resultados pode ser verificada através de estudos de simulação (Ratkowsky, 1983); pelas medidas de curvatura (Bates e Watts, 1980); ou através das medidas de vícios dos estimadores (Box, 1971). Em muitos casos, a precisão dos resultados aproximados pode ser melhorada considerando uma reparametrização apropriada. Uma outra alternativa é a obtenção das inferências sobre os parâmetros do modelo não-linear através da análise Bayesiana.

Frequentemente os estatísticos e pesquisadores fazem inferências sobre funções de parâmetros e geralmente nessas situações, não é possível obter resultados exatos, sendo comum considerar-se a normalidade assintótica dos estimadores de máxima verossimilhança, baseada na informação de Fisher ou na informação observada (ver por exemplo, Bickel e Doksum, 1977). Esses resultados podem não ser apropriados quando trabalhamos com amostras pequenas ou moderadas. Assim, uma alternativa é considerar reparametrizações que melhorem a normalidade da função de verossimilhança (Sprott, 1973).

Em nosso trabalho, vamos explorar duas transformações paramétricas sugeridas na literatura para o modelo de Michaelis-Menten: uma reparametrização proposta por Bates e Watts (1988) e uma proposta por Seber e Wild (1989), que serão apresentadas a seguir.

3.1. Algumas Reparametrizações para o Modelo de Michaelis-Menten

Nesta seção, com o propósito de verificar se a utilização de diferentes reparametrizações proporcionam melhores resultados em termos de aproximações na obtenção das densidades a posteriori ou apresentam alguma melhoria na normalidade da função de verossimilhança, vamos considerar duas reparametrizações para o modelo de Michaelis-Menten propostas na literatura.

3.1.1. Reparametrização proposta por Bates e Watts

Uma reparametrização proposta por Bates e Watts (1988) é definida na forma,

$$\begin{cases} \beta_1 = \frac{\theta_2}{\theta_1} \\ \beta_2 = \frac{1}{\theta_1} \end{cases}, \text{ isto é, } \begin{cases} \theta_1 = \frac{\beta_1}{\beta_2} \\ \theta_2 = \frac{1}{\beta_2} \end{cases} \quad (3.1)$$

com $0 < \beta_1, \beta_2 < \infty$.

3.1.2. Reparametrização proposta por Seber e Wild

Uma outra reparametrização sugerida na literatura para o modelo não-linear de Michaelis-Menten é proposta por Seber e Wild (1989), definida por,

$$\begin{cases} \delta_1 = \frac{\theta_1 u}{\theta_2 + u} \\ \delta_2 = \frac{\theta_1 v}{\theta_2 + v} \end{cases}, \text{ isto é, } \begin{cases} \theta_1 = \frac{\delta_1 \delta_2 (u - v)}{(u \delta_2 - v \delta_1)} \\ \theta_2 = \frac{uv \cdot (\delta_1 - \delta_2)}{(u \delta_2 - v \delta_1)} \end{cases} \quad (3.2)$$

onde u e v são valores apropriados para a variável independente x . Para encontrar os valores apropriados para u e v na transformação (3.2), podemos considerar gráficos de contorno para a função de verossimilhança de (δ_1, δ_2) , assumindo σ^2 conhecido. Assim, obtemos *boa normalidade* para a função de verossimilhança se os contornos apresentarem formas elípticas. Kass e Slate (1992) e Hills e Smith (1993) propõem medidas e diagnósticos para verificar se uma dada reparametrização é viável.

Caso os parâmetros originais tenham interpretações físicas importantes, nas quais o pesquisador tiver interesse, podemos estimar os parâmetros através de uma análise Bayesiana.

As duas reparametrizações apresentadas, podem ser utilizadas na análise clássica do modelo de Michaelis-Menten e também na análise Bayesiana, especificamente para *melhorar* a precisão dos resultados Bayesianos aproximados (ver Achcar e Smith, 1990).

3.2. Medidas e Diagnósticos para a Normalidade

Ao adotarmos a normalidade assintótica para o estimador de máxima verossimilhança ou os resultados da teoria clássica apresentados no Capítulo 2, podemos estar assumindo uma aproximação que pode não condizer com a realidade, assim, torna-se necessária a verificação dessa suposta normalidade.

É muito importante verificar o quanto conseguimos melhorar a normalidade da função de verossimilhança ou de densidades a posteriori de interesse, quando utilizamos uma reparametrização.

Assim, uma forma bem simples para se verificar a normalidade da função de verossimilhança ou de densidades a posteriori, é através de gráficos e contornos. No caso bivariado, por exemplo, aceitamos a normalidade quando encontramos contornos com formas de elipses aproximadas.

Quando temos interesse em um dado parâmetro e existem vários parâmetros auxiliares (nuisance parameters), podemos, como uma alternativa, considerar a verossimilhança *profile* e assim, verificar sua normalidade através do método gráfico proposto por Hills e Smith.

3.2.1. T-Plot de Hills e Smith

Ao realizar uma análise gráfica da função de verossimilhança, geralmente, ficamos restritos a avaliar apenas a simetria ou elipticidade dos gráficos, não tendo como medir o seu *grau* de normalidade, ou seja, o quanto estamos próximos ou não, da normalidade desejada.

Hills e Smith (1993) introduziram um método gráfico bastante prático para se verificar a normalidade de funções de verossimilhança ou densidades a posteriori, o método T-Plot, como generalização e extensão a um método proposto por Bates e Watts (1980), onde a normalidade pode ser avaliada através da linearidade dos gráficos.

Seja, $p(\hat{\theta} / \text{dados})$ o valor da densidade a posteriori calculada na moda $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$ e seja $p(\theta_j / \text{dados})$ a densidade a posteriori *profile* de θ_j , dada por,

$$p(\theta_j / \text{dados}) = \max_{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p} p(\theta / \text{dados})$$

Uma medida de diagnóstico é dada por,

$$T(\theta_j) = \text{sgn}(\theta_j - \hat{\theta}_j) \left\{ -2 \ln \left[\frac{p(\theta_j / \text{dados})}{p(\hat{\theta}_j / \text{dados})} \right] \right\}^{\frac{1}{2}} \quad (3.3)$$

onde $\text{sgn}(x) = x/|x|$ é a função sinal de x .

O T-Plot é definido como o gráfico de $T(\theta_j)$ versus θ_j , que deverá ser uma linha reta caso o parâmetro θ_j seja normalmente distribuído, conseqüentemente o gráfico pode ser usado para identificar desvios da normalidade.

O T-Plot de Hills e Smith, no caso *unidimensional*, é, simplesmente, definido por,

$$T(\theta) = \text{sgn}(\theta - \hat{\theta}) \left\{ -2 \ln \left[\frac{p(\theta / \text{dados})}{p(\hat{\theta} / \text{dados})} \right] \right\}^{\frac{1}{2}} \quad (3.4)$$

O caso *bidimensional* é mais complexo, pois a extensão óbvia de (3.4) não funciona. Nesse caso, aplicamos a expansão de Taylor em $\log[p(\theta_1, \theta_2 / \text{dados})]$ em torno de $\hat{\theta}$, a moda de $p(\theta_1, \theta_2 / \text{dados})$. Assim, procedendo como no caso anterior, construímos o contorno de $T(\theta_1, \theta_2)$ versus (θ_1, θ_2) . Caso (θ_1, θ_2) sejam normalmente distribuídos, então o contorno deverá ter forma elíptica. Entretanto, surge um problema: é mais difícil julgar o *grau de elipticidade* do que o *grau de linearidade*; assim, podemos simplesmente construir contornos para a densidade a posteriori $p(\theta_1, \theta_2 / \text{dados})$ e analisá-los diretamente. A solução, então, seria fixarmos um parâmetro e analisar o comportamento do outro parâmetro isoladamente.

Utilizaremos, em nosso trabalho, o T-Plot proposto por Hills e Smith para a função de verossimilhança $L(\theta)$, a partir da seguinte relação,

$$\begin{aligned} T(\theta) &= \text{sgn}(\theta - \hat{\theta}) \left\{ -2 \ln \left[\frac{L(\theta)}{L(\hat{\theta})} \right] \right\}^{\frac{1}{2}} \\ &= \text{sgn}(\theta - \hat{\theta}) \left\{ -2 [l(\theta) - l(\hat{\theta})] \right\}^{\frac{1}{2}} \end{aligned} \quad (3.5)$$

onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ e $l(\theta)$ é o logaritmo da função de verossimilhança.

3.2.2. Terceira Derivada Padronizada do Logaritmo da Verossimilhança

Kass e Slate (1992) propõem uma medida de não-normalidade de densidades a posteriori ou de funções de verossimilhança, dada pela terceira derivada padronizada, STD , que simplifica as interpretações, pois reduz as análises a um simples escalar. Originalmente, a terceira derivada padronizada do logaritmo da verossimilhança foi proposta por Sprott (1973) e é definida por,

$$STD(\hat{\theta}) = \left| l''(\hat{\theta}) \cdot [l''(\hat{\theta})]^{-\frac{3}{2}} \right| \quad (3.6)$$

onde $l''(\hat{\theta})$ e $l'''(\hat{\theta})$ são, respectivamente, a segunda e terceira derivadas do logaritmo da função de verossimilhança calculadas no estimador de máxima verossimilhança $\hat{\theta}$. Caso a função de verossimilhança apresente *boa normalidade*, esperamos encontrar $STD(\hat{\theta})$, dado em (3.6), próximo ou igual a zero.

Analogamente, podemos calcular a terceira derivada padronizada para a densidade a posteriori, $p(\theta / \text{dados})$, simplesmente substituindo o estimador de máxima verossimilhança $\hat{\theta}$ pela moda a posteriori $\tilde{\theta}$.

Uma generalização da terceira derivada padronizada, no caso multiparamétrico, é dada pela medida global de não-normalidade de Kass e Slate (1992), dada por,

$$m^2 \bar{B}^2(\hat{\theta}) = \sum_{a,b,c,d,e,f} d_{abc} d_{def} b_{ab} b_{de} b_{cf} \quad (3.7)$$

onde $d_{abc} = \frac{\partial^3 l(\hat{\theta})}{\partial \theta_a \partial \theta_b \partial \theta_c}$; b_{ab} são os elementos da inversa da matriz de informação e m é o número de parâmetros.

Se os parâmetros considerados tiverem uma *boa normalidade* conjunta, esperamos encontrar $m^2 \bar{B}^2(\hat{\theta}) \approx 0$.

Devemos observar, que a medida global dada em (3.7) reduz-se à (3.6) no caso uniparamétrico.

Em nosso trabalho, utilizaremos a medida global de não-normalidade $m^2 \bar{B}^2$, para verificar a normalidade da função de verossimilhança para o modelo de Michaelis-Menten, utilizando as reparametrizações mencionadas.

3.3. Inferências para o Modelo de Michaelis-Menten considerando as Reparametrizações Propostas

Como visto anteriormente, os resultados assintóticos da teoria clássica apresentados no capítulo anterior, podem não ser precisos quando trabalhamos com pequenas amostras ou quando o modelo considerado for altamente não-linear. Assim, a eficiência desses resultados pode ser verificada através de simulação (ver Apêndice C), pelas medidas de curvatura (Bates e Watts, 1980) ou através de medidas de vícios (ver por exemplo, Box, 1971). Em muitos casos, a precisão dos resultados aproximados pode ser melhorada considerando uma reparametrização apropriada.

Caminhos alternativos para verificar uma dada reparametrização são propostos por Kass e Slate (1992) e Hills e Smith (1993).

Nesta seção, vamos apresentar inferências, baseadas na normalidade assintótica dos estimadores de máxima verossimilhança, para os parâmetros do modelo, considerando inicialmente a reparametrização proposta por Bates e Watts, dada em (3.1). A função de verossimilhança para β_1, β_2 e σ^2 é dada por,

$$L(\beta_1, \beta_2, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{x_i}{\beta_1 + \beta_2 x_i}\right)^2\right\} \quad (3.8)$$

onde $0 < \beta_1, \beta_2 < \infty$ e $\sigma^2 > 0$.

O logaritmo da função de verossimilhança (3.8) é dado por,

$$l(\beta_1, \beta_2, \sigma^2) \propto -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{x_i}{\beta_1 + \beta_2 x_i}\right)^2$$

Os estimadores de máxima verossimilhança para os parâmetros β_1, β_2 e σ^2 são obtidos através das derivadas parciais,

$$\begin{cases} \frac{\partial l(\beta_1, \beta_2, \sigma^2)}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i [y_i \beta_1 + x_i (y_i \beta_2 - 1)]}{(\beta_1 + \beta_2 x_i)^3} \right\} \\ \frac{\partial l(\beta_1, \beta_2, \sigma^2)}{\partial \beta_2} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i^2 [y_i \beta_1 + x_i (y_i \beta_2 - 1)]}{(\beta_1 + \beta_2 x_i)^3} \right\} \\ \frac{\partial l(\beta_1, \beta_2, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{S(\beta_1, \beta_2)}{2(\sigma^2)^2} \end{cases}$$

onde $S(\beta_1, \beta_2) = \sum_{i=1}^n \left(y_i - \frac{x_i}{\beta_1 + \beta_2 x_i} \right)^2$.

Utilizando, novamente, um método iterativo, por exemplo, Newton-Raphson, encontramos $\hat{\beta}_1, \hat{\beta}_2$ e $\hat{\sigma}^2$, tal que: $\frac{\partial l}{\partial \beta_1} = 0, \frac{\partial l}{\partial \beta_2} = 0$ e $\frac{\partial l}{\partial \sigma^2} = 0$, sendo que $\hat{\sigma}^2$ é dado por,

$$\hat{\sigma}^2 = \frac{S(\hat{\beta}_1, \hat{\beta}_2)}{n}$$

Os estimadores de máxima verossimilhança $\hat{\beta}_1, \hat{\beta}_2$ e $\hat{\sigma}^2$ têm normalidade assintótica dada por,

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) \overset{a}{\sim} N_3\{(\beta_1, \beta_2, \sigma^2); I^{-1}(\beta_1, \beta_2, \sigma^2)\} \quad (3.9)$$

onde $I(\beta_1, \beta_2, \sigma^2)$ é a matriz de informação de Fisher, dada por,

$$I(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^2} \begin{pmatrix} \sum_{i=1}^n \frac{x_i^2}{(\beta_1 + \beta_2 x_i)^4} & \sum_{i=1}^n \frac{x_i^3}{(\beta_1 + \beta_2 x_i)^4} & 0 \\ \sum_{i=1}^n \frac{x_i^3}{(\beta_1 + \beta_2 x_i)^4} & \sum_{i=1}^n \frac{x_i^4}{(\beta_1 + \beta_2 x_i)^4} & 0 \\ 0 & 0 & \frac{n}{2\sigma^2} \end{pmatrix}$$

Em problemas práticos, também é usual considerar a matriz de informação observada em lugar da matriz de informação de Fisher.

Podemos construir intervalos de confiança para os parâmetros β_1, β_2 e σ^2 , considerando a normalidade assintótica dos estimadores de máxima verossimilhança dada em (3.9).

Considerando a reparametrização proposta por Seber e Wild (1989), dada em (3.2), temos que a função de verossimilhança para δ_1, δ_2 e σ^2 é dada por,

$$L(\delta_1, \delta_2, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left[y_i - \frac{(u-v)\delta_1\delta_2 x_i}{[v(u-x_i)\delta_1 - u(v-x_i)\delta_2]} \right]^2 \right\} \quad (3.10)$$

onde $0 < \delta_1, \delta_2 < \infty$ e $\sigma^2 > 0$.

O logaritmo da função de verossimilhança (3.10) é dado por,

$$l(\delta_1, \delta_2, \sigma^2) \propto -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left\{ y_i - \frac{\delta_1\delta_2 a_i}{(\delta_1 b_i - \delta_2 c_i)} \right\}^2$$

onde

$$\begin{aligned} a_i &= (u-v)x_i, \\ b_i &= v(u-x_i), \\ c_i &= u(v-x_i). \end{aligned}$$

Os estimadores de máxima verossimilhança para os parâmetros δ_1, δ_2 e σ^2 , são obtidos através das derivadas parciais,

$$\begin{aligned}
\frac{\partial l(\delta_1, \delta_2, \sigma^2)}{\partial \delta_1} &= \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \left(y_i - \frac{a_i \delta_1 \delta_2}{\delta_1 b_i - \delta_2 c_i} \right) \cdot \left(-\frac{a_i c_i \delta_1 \delta_2}{\delta_1^2 (\delta_1 b_i - \delta_2 c_i)^2} \right) \right\} \\
\frac{\partial l(\delta_1, \delta_2, \sigma^2)}{\partial \delta_2} &= \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \left(y_i - \frac{a_i \delta_1 \delta_2}{\delta_1 b_i - \delta_2 c_i} \right) \cdot \left(\frac{a_i b_i \delta_1^2 \delta_2^2}{\delta_2^2 (\delta_1 b_i - \delta_2 c_i)^2} \right) \right\} \\
\frac{\partial l(\delta_1, \delta_2, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \left\{ y_i - \frac{a_i \delta_1 \delta_2}{\delta_1 b_i - \delta_2 c_i} \right\}^2
\end{aligned} \tag{3.11}$$

Devemos utilizar um método iterativo, por exemplo, Newton-Raphson, para calcular $\hat{\delta}_1, \hat{\delta}_2$ e $\hat{\sigma}^2$, tal que: $\partial l / \partial \delta_1 = 0$, $\partial l / \partial \delta_2 = 0$ e $\partial l / \partial \sigma^2 = 0$, sendo que $\hat{\sigma}^2$ é dado por,

$$\hat{\sigma}^2 = \frac{S(\hat{\delta}_1, \hat{\delta}_2)}{n}$$

onde $S(\delta_1, \delta_2) = \sum_{i=1}^n \left\{ y_i - \frac{a_i \delta_1 \delta_2}{\delta_1 b_i - \delta_2 c_i} \right\}^2$. Assim, reduzimos (3.11) a um sistema de duas equações.

Os e.m.v. $\hat{\delta}_1, \hat{\delta}_2$ e $\hat{\sigma}^2$ têm normalidade assintótica dada por,

$$(\hat{\delta}_1, \hat{\delta}_2, \hat{\sigma}^2) \stackrel{a}{\sim} N_3[(\delta_1, \delta_2, \sigma^2); I^{-1}(\delta_1, \delta_2, \sigma^2)] \tag{3.12}$$

onde $I(\delta_1, \delta_2, \sigma^2)$ é a matriz de informação de Fisher dada por,

$$I(\delta_1, \delta_2, \sigma^2) = \frac{1}{\sigma^2} \begin{pmatrix} \delta_2^4 \sum_{i=1}^n \left(\frac{a_i^2 c_i^2}{(\delta_1 b_i - \delta_2 c_i)^4} \right) & -\delta_1^2 \delta_2^2 \sum_{i=1}^n \left(\frac{a_i^2 b_i c_i}{(\delta_1 b_i - \delta_2 c_i)^4} \right) & 0 \\ -\delta_1^2 \delta_2^2 \sum_{i=1}^n \left(\frac{a_i^2 b_i c_i}{(\delta_1 b_i - \delta_2 c_i)^4} \right) & \delta_1^4 \sum_{i=1}^n \left(\frac{a_i^2 b_i^2}{(\delta_1 b_i - \delta_2 c_i)^4} \right) & 0 \\ 0 & 0 & \frac{n}{2\sigma^2} \end{pmatrix}$$

Em geral, devemos tomar cuidado com o uso desses resultados assintóticos, especialmente para amostras pequenas ou moderadas, pois as aproximações (3.9) e (3.12) podem não ser boas, comprometendo a precisão das inferências. Para tanto, necessitamos de uma reparametrização que forneça boa normalidade para a função de verossimilhança. Assim, para avaliar a normalidade das funções de verossimilhança apresentadas podemos utilizar a medida global de Kass e Slate e o *T-Plot* de Hills e Smith.

3.4. Inferências para o Modelo de Michaelis-Menten considerando a Função de Verossimilhança *Profile*

Com θ_2 desconhecido, as inferências sobre θ_1 são mais complexas, uma vez que devemos estudar a normalidade conjunta de dois parâmetros do nosso modelo. Para amenizar nossa dificuldade, vamos considerar, como alternativa, a função de verossimilhança *profile* para θ_1 .

Desta forma, considerando θ_2 desconhecido, podemos calcular a função de verossimilhança *profile* para θ_1 , $L_{\hat{\theta}_2}(\theta_1) = L(\theta_1, \hat{\theta}_2)$, onde $\hat{\theta}_2$ é o valor que maximiza o logaritmo da função de verossimilhança para cada valor de θ_1 . Portanto, o logaritmo da verossimilhança *profile* é dado por,

$$l_{\hat{\theta}_2}(\theta_1) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\hat{\theta}_2 + x_i} \right)^2 \quad (3.13)$$

onde $\hat{\theta}_2$ é o valor que maximiza o logaritmo da função de verossimilhança dado em (3.13), para cada valor de θ_1 e σ^2 é considerado constante.

• *Diagnósticos e Medidas para a Normalidade:*

i) T-Plot de Hills e Smith

O T-Plot de Hills e Smith deve ser obtido ponto a ponto através da relação,

$$T(\theta_1) = \text{sgn}(\theta_1 - \hat{\theta}_1) \left\{ -2 \left[l_{\hat{\theta}_2}(\theta_1) - l(\hat{\theta}_1, \hat{\theta}_2) \right] \right\} \quad (3.14)$$

onde $\hat{\theta}_2$ maximiza (3.14) para cada valor de θ_1 .

Analogamente, podemos obter o T-Plot de Hills e Smith para o logaritmo da função de verossimilhança *profile* de θ_2 pela relação,

$$T(\theta_2) = \text{sgn}(\theta_2 - \hat{\theta}_2) \left\{ -2 \left[l_{\hat{\theta}_1}(\theta_2) - l(\hat{\theta}_1, \hat{\theta}_2) \right] \right\} \quad (3.15)$$

onde $\hat{\theta}_1$ maximiza (3.15) para cada valor de θ_2 .

ii) Medida global de não-normalidade de Kass e Slate

Também temos interesse em utilizar a medida global de não-normalidade de Kass e Slate, definida em (3.7). As derivadas terceiras parciais de $l(\theta_1, \theta_2)$, em relação a θ_1 e θ_2 , calculadas no ponto de máxima verossimilhança, são dadas por,

$$\begin{aligned} \frac{\partial^3 l(\theta_1, \theta_2)}{\partial \theta_1^2 \partial \theta_2} \Big|_{\hat{\theta}_1, \hat{\theta}_2} &= \frac{2}{\sigma^2} \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^4} \\ \frac{\partial^3 l(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2^2} \Big|_{\hat{\theta}_1, \hat{\theta}_2} &= \frac{2}{\sigma^2} \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^3} - \frac{6\hat{\theta}_1}{\sigma^2} \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^4} \end{aligned} \quad (3.16)$$

$$\frac{\partial^3 l(\theta_1, \theta_2)}{\partial \theta_1^3} \Big|_{\hat{\theta}_1, \hat{\theta}_2} = 0$$

$$\frac{\partial^3 l(\theta_1, \theta_2)}{\partial \theta_2^3} \Big|_{\hat{\theta}_1, \hat{\theta}_2} = -\frac{6\hat{\theta}_1}{\sigma^2} \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^4} + \frac{12\hat{\theta}_1^2}{\sigma^2} \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^5}$$

onde σ^2 é considerado conhecido.

Dessa forma, aplicando a medida $m^2 \bar{B}^2$ podemos avaliar a normalidade conjunta dos estimadores de máxima verossimilhança na parametrização original (θ_1, θ_2) .

3.4.1. Reparametrização proposta por Bates e Watts para a Função de Verossimilhança *Profile*

Considerando a reparametrização proposta por Bates e Watts (1988), dada em (3.2), o logaritmo da função de verossimilhança *profile* para β_1 é dado por,

$$l_{\hat{\beta}_2}(\beta_1) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left\{ y_i - \frac{x_i}{(\beta_1 + \hat{\beta}_2 x_i)} \right\}^2 \quad (3.17)$$

onde $\hat{\beta}_2$ é o estimador de máxima verossimilhança do logaritmo da função de verossimilhança para cada valor de β_1 e σ^2 é considerado conhecido.

• Medidas e diagnósticos para a Normalidade:

i) T-Plot de Hills e Smith

O T-Plot de Hills e Smith para a função de verossimilhança *profile* de β_1 também é obtido ponto a ponto através da relação,

$$T(\beta_1) = \text{sgn}(\beta_1 - \hat{\beta}_1) \left\{ -2 \left[l_{\hat{\beta}_2}(\beta_1) - l(\hat{\beta}_1, \hat{\beta}_2) \right] \right\} \quad (3.18)$$

onde $\hat{\beta}_2$ maximiza (3.18) para cada valor de β_1 .

Similarmente, o *T-Plot* de Hills e Smith para a função de verossimilhança *profile* de β_2 deve ser obtido ponto a ponto pela relação,

$$T(\beta_2) = \text{sgn}(\beta_2 - \hat{\beta}_2) \left\{ -2 \left[l_{\hat{\beta}_1}(\beta_2) - l(\hat{\beta}_1, \hat{\beta}_2) \right] \right\} \quad (3.19)$$

onde $\hat{\beta}_1$ maximiza (3.19) para cada valor de β_2 .

ii) Medida global de não-normalidade de Kass e Slate

Considerando a reparametrização $(\hat{\beta}_1, \hat{\beta}_2)$, podemos utilizar a medida global de não-normalidade de Kass e Slate (3.7). As derivadas terceiras parciais de $l(\beta_1, \beta_2)$, em relação a β_1 e β_2 , calculadas nos estimadores de máxima verossimilhança, são dadas por,

$$\begin{aligned} \frac{\partial^3 l(\beta_1, \beta_2)}{\partial \beta_1^3} \Big|_{\hat{\beta}_1, \hat{\beta}_2} &= -\frac{6}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i [y_i \hat{\beta}_1 + x_i (y_i \hat{\beta}_2 - 2)]}{(\hat{\beta}_1 + \hat{\beta}_2 x_i)^5} \right\} \\ \frac{\partial^3 l(\beta_1, \beta_2)}{\partial \beta_2^3} \Big|_{\hat{\beta}_1, \hat{\beta}_2} &= -\frac{6}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i^4 [y_i \hat{\beta}_1 + x_i (y_i \hat{\beta}_2 - 2)]}{(\hat{\beta}_1 + \hat{\beta}_2 x_i)^5} \right\} \\ \frac{\partial^3 l(\beta_1, \beta_2)}{\partial \beta_1^2 \partial \beta_2} \Big|_{\hat{\beta}_1, \hat{\beta}_2} &= -\frac{6}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i^2 [y_i \hat{\beta}_1 + x_i (y_i \hat{\beta}_2 - 2)]}{(\hat{\beta}_1 + \hat{\beta}_2 x_i)^5} \right\} \\ \frac{\partial^3 l(\beta_1, \beta_2)}{\partial \beta_1 \partial \beta_2^2} \Big|_{\hat{\beta}_1, \hat{\beta}_2} &= -\frac{6}{\sigma^2} \sum_{i=1}^n \left\{ \frac{x_i^3 [y_i \hat{\beta}_1 + x_i (y_i \hat{\beta}_2 - 2)]}{(\hat{\beta}_1 + \hat{\beta}_2 x_i)^5} \right\} \end{aligned} \quad (3.20)$$

onde σ^2 é considerado conhecido.

3.4.2. Reparametrização proposta por Seber e Wild para a Função de Verossimilhança *Profile*

Nesta subseção, considerando a reparametrização proposta por Seber e Wild (1989), dada em (3.2), o logaritmo da função de verossimilhança *profile* para δ_1 é dado por,

$$l_{\hat{\delta}_2}(\delta_1) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left\{ y_i - \frac{(u-v)\delta_1\hat{\delta}_2x_i}{[v(u-x_i)\delta_1 + u(x_i-v)\hat{\delta}_2]} \right\}^2 \quad (3.21)$$

onde σ^2 é considerado conhecido e para cada valor de δ_1 , $\hat{\delta}_2$ é o estimador que maximiza o logaritmo da função de verossimilhança dado em (3.21).

• Medidas e diagnósticos para a Normalidade:

i) T-Plot de Hills e Smith

O *T-Plot* de Hills e Smith deve ser obtido ponto a ponto pela relação,

$$T(\delta_1) = \text{sgn}(\delta_1 - \hat{\delta}_1) \left\{ -2 \left[l_{\hat{\delta}_2}(\delta_1) - l(\hat{\delta}_1, \hat{\delta}_2) \right] \right\} \quad (3.22)$$

onde $\hat{\delta}_2$ maximiza (3.22) para cada valor de δ_1 .

Analogamente, podemos obter o *T-Plot* de Hills e Smith para o logaritmo da função de verossimilhança *profile* de δ_2 pela relação,

$$T(\delta_2) = \text{sgn}(\delta_2 - \hat{\delta}_2) \left\{ -2 \left[l_{\hat{\delta}_1}(\delta_2) - l(\hat{\delta}_1, \hat{\delta}_2) \right] \right\} \quad (3.23)$$

onde $\hat{\delta}_1$ maximiza (3.23) para cada valor de δ_2 .

ii) Medida global de não-normalidade de Kass e Slate

Para avaliar a normalidade conjunta dos estimadores de máxima verossimilhança na parametrização $(\hat{\delta}_1, \hat{\delta}_2)$, podemos utilizar a medida global de não-normalidade de Kass e Slate definida em (3.7). As derivadas parciais de $l(\delta_1, \delta_2)$, em relação a δ_1 e δ_2 , calculadas no ponto de máxima verossimilhança são dadas por,

$$\begin{aligned} \frac{\partial^3 l(\delta_1, \delta_2)}{\partial \delta_1^3} \Big|_{\hat{\delta}_1, \hat{\delta}_2} &= 6uv\hat{\delta}_2^2(u-v) \times \sum_{i=1}^n x_i \cdot (x_i - v) \cdot (x_i - u) \times \\ &\times \left\{ \frac{v\hat{\delta}_1(x_i - u)[\hat{\delta}_2 x_i(u-v) + y_i v(x_i - u)] + u\hat{\delta}_2(x_i - v)[\hat{\delta}_2 x_i(u-v) - y_i v(x_i - u)]}{\sigma^2 [v\hat{\delta}_1(x_i - u) - u\hat{\delta}_2(x_i - v)]^5} \right\} \\ \frac{\partial^3 l(\delta_1, \delta_2)}{\partial \delta_2^3} \Big|_{\hat{\delta}_1, \hat{\delta}_2} &= -6uv\hat{\delta}_1^2(u-v) \times \\ &\times \sum_{i=1}^n x_i(x_i - v) \cdot (x_i - u) \left\{ \frac{v\hat{\delta}_1^2 x_i(u-v)(x_i - u) + u\hat{\delta}_1(x_i - v)[\hat{\delta}_2 x_i(u-v) + y_i v(x_i - u)] - y_i u^2 \hat{\delta}_2(x_i - v)^2}{\sigma^2 [v\hat{\delta}_1(x_i - u) - u\hat{\delta}_2(x_i - v)]^5} \right\} \\ \frac{\partial^3 l(\delta_1, \delta_2)}{\partial \delta_1^2 \partial \delta_2} \Big|_{\hat{\delta}_1, \hat{\delta}_2} &= -2uv\hat{\delta}_2(u-v) \sum_{i=1}^n x_i(x_i - v) \cdot (x_i - u) \times \\ &\times \left\{ \frac{v\hat{\delta}_1^2(x_i - u)[3\hat{\delta}_2 x_i(u-v) + 2y_i v(x_i - u)] + u\hat{\delta}_1 \hat{\delta}_2(x_i - v)[3\hat{\delta}_2 x_i(u-v) - y_i v(x_i - u)] - y_i u^2 \hat{\delta}_2^2(x_i - v)^2}{\sigma^2 [v\hat{\delta}_1(x_i - u) - u\hat{\delta}_2(x_i - v)]^5} \right\} \\ \frac{\partial^3 l(\delta_1, \delta_2)}{\partial \delta_1 \partial \delta_2^2} \Big|_{\hat{\delta}_1, \hat{\delta}_2} &= 2uv\hat{\delta}_1(u-v) \sum_{i=1}^n x_i(x_i - v) \cdot (x_i - u) \times \\ &\times \left\{ \frac{v\hat{\delta}_1^2(x_i - u)[3\hat{\delta}_2 x_i(u-v) + y_i v(x_i - u)] - 2y_i u^2 \hat{\delta}_2^2(x_i - v)^2}{\sigma^2 [v\hat{\delta}_1(x_i - u) - u\hat{\delta}_2(x_i - v)]^5} \right\} \end{aligned} \quad (3.24)$$

onde σ^2 é considerado conhecido.

No capítulo 7, podemos verificar, com os conjuntos de dados, a melhoria significativa que essas reparametrizações proporcionam na normalidade da função de verossimilhança para o modelo de Michaelis-Menten, especialmente a reparametrização proposta por Seber e Wild (3.2).

Capítulo 4

ESTIMAÇÃO BAYESIANA PARA MODELOS NÃO-LINEARES

Em análise de regressão não-linear, temos interesse em estimar a relação existente entre duas ou mais variáveis aleatórias, através de um modelo e determinar o valor dos parâmetros que o definem.

Embora os parâmetros são considerados fixos, não sabemos o seu verdadeiro valor e uma forma de expressar nossa incerteza, antes de relacionar uma amostra, é através de uma distribuição de probabilidade a priori para os parâmetros do modelo, que sob este ponto de vista é considerado uma variável aleatória. O *teorema de Bayes* (ver por exemplo, Box e Tiao, 1973) fornece um método para calcular a densidade de probabilidade de θ , incorporando a informação dos dados observados, com a informação adicional dada pela densidade a priori.

Inicialmente, para representar a opinião prévia sobre os parâmetros, vamos considerar uma densidade a priori não-informativa para os parâmetros do modelo e posteriormente densidades a priori informativas, para assim, obter as densidades a posteriori de interesse. Vamos explorar o uso de algumas reparametrizações para o modelo de Michaelis-Menten, sugeridas na literatura, para *melhorar* a normalidade das densidades a posteriori de interesse.

4.1. Fórmula de Bayes

Sejam Y e θ variáveis aleatórias com densidades conjunta e condicionais dadas por $p(y, \theta)$, $p(y / \theta)$ e $p(\theta / y)$, respectivamente e seja $p(\theta)$ a densidade marginal de θ . A densidade condicional de θ dado y é dada por,

$$p(\theta / y) = \frac{p(y, \theta)}{p(y)}$$

$$p(\theta / y) = \frac{p(y / \theta)p(\theta)}{\int_{\theta \in \Theta} p(y / \theta)p(\theta)d\theta}, \quad (4.1)$$

onde Θ é o espaço paramétrico.

4.2. Densidades a Priori para os Parâmetros

Antes de analisar estatisticamente os dados, devemos incorporar o nosso conhecimento prévio a respeito dos parâmetros do modelo. Uma maneira conveniente de se fazer isto é formular uma densidade de probabilidade para os parâmetros θ . Essa densidade de probabilidade para θ é identificada como *densidade a priori* e é denotada por $p(\theta)$. Quando não temos nenhuma informação a respeito de θ , devemos formular uma priori, tal que toda informação acerca do parâmetro seja exclusivamente dos dados, ou seja, devemos considerar uma *priori não-informativa*. Várias regras para a formulação de uma priori não-informativa estão definidas, entre as quais podemos destacar a regra de Bernardo, Jeffreys (1961)¹ e Zellner.

Neste trabalho, destacamos o uso da densidade a priori não-informativa de Jeffreys (ver por exemplo, Box e Tiao, 1973).

Seja $l(\theta)$ o logaritmo da função de verossimilhança para o parâmetro θ de um modelo; a densidade a priori de Jeffreys para θ é dada por,

$$p(\theta) \propto \{\det(\mathbf{I}(\theta))\}^{\frac{1}{2}}, \quad (4.2)$$

onde $\mathbf{I}(\theta)$ é a matriz de informação esperada (Fisher), dada por,

$$E\left[-\frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s}\right]. \quad (4.3)$$

Quando $\theta = (\alpha', \beta')'$ e supondo independência a priori, isto é, $p(\alpha, \beta) = p_1(\alpha)p_2(\beta)$, para obter $p_1(\alpha)$ e $p_2(\beta)$ separadamente, a regra de Jeffreys trata os outros parâmetros como conhecidos. Por exemplo,

$$p_1(\alpha) \propto \{\det(\mathbf{I}_1(\alpha))\}^{1/2} \quad (4.4)$$

onde,

$$[\mathbf{I}_1(\alpha)]_{r,s} = E\left[-\frac{\partial^2 l(\alpha, \beta)}{\partial \alpha_r \partial \alpha_s}\right].$$

¹A regra de Jeffreys para a construção de priori não-informativas é discutida por Box e Tiao, (1973).

Aplicando a idéia acima ao modelo não-linear geral (1.10) dado na forma matricial,

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \quad (4.5)$$

onde $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$; temos de (1.50),

$$l(\boldsymbol{\theta}, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} S(\boldsymbol{\theta}),$$

onde $S(\boldsymbol{\theta}) = \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|^2$. Portanto, de (1.46), temos,

$$\begin{aligned} \mathbf{I}_1(\boldsymbol{\theta}) &= \frac{1}{2\sigma^2} \mathcal{E} \left(\frac{\partial^2 S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \\ &= \frac{1}{\sigma^2} \mathbf{F}'(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}), \\ &\propto \mathbf{F}'(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}) \end{aligned}$$

e

$$I_2(\sigma) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4} E[S(\boldsymbol{\theta})] = \frac{2n}{\sigma^2}.$$

Assim, uma priori não-informativa para $(\boldsymbol{\theta}, \sigma)$, onde $\boldsymbol{\theta}$ e σ são considerados independentes a priori, é dada por,

$$p(\boldsymbol{\theta}, \sigma) \propto \det(\mathbf{F}'(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}))^{1/2} \sigma^{-1}. \quad (4.6)$$

A priori (4.6) acima também foi utilizada por Eaves (1983).

No caso onde temos um modelo linear (ver por exemplo Seber, 1977) $\mathbf{F}'(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}) = \mathbf{X}'\mathbf{X}$, é independente de $\boldsymbol{\theta}$, tal que $p_1(\boldsymbol{\theta})$ é proporcional a uma constante e portanto,

$$p(\boldsymbol{\theta}, \sigma) \propto \sigma^{-1}. \quad (4.7)$$

Por analogia com o caso linear, Box e Tiao (1973: p. 431) argumentam que se existe uma parametrização do modelo não-linear apropriada, então devemos considerar (4.7) como uma razoável aproximação da priori não-informativa dada em (4.6). A priori (4.7) foi discutida por Zellner (1971), mas o seu uso indiscriminado nos problemas em que existe

grande não-linearidade intrínseca ou grande não-linearidade paramétrica, essa priori torna-se uma péssima aproximação para a priori não-informativa (4.6).

4.3. Densidades a Posteriori para os Parâmetros

Uma vez definidos o modelo para os dados e a distribuição a priori, devemos combinar a informação prévia sobre o parâmetro θ com a informação contida nos dados, obtendo assim uma *distribuição a posteriori* para θ .

Se $p(\theta, \sigma)$ representa a distribuição a priori dos parâmetros desconhecidos e $p(y/\theta, \sigma)$ é a função de verossimilhança, freqüentemente, escrita como $L(\theta, \sigma)$, podemos obter as distribuições a posteriori conjuntas e marginais para os parâmetros. Considerando o modelo não-linear dado na forma matricial (4.5), onde $\varepsilon \sim N_n(0, \sigma^2 \mathbf{I}_n)$, a função densidade a posteriori para θ e σ , pela fórmula de Bayes (4.1), é dada por,

$$p(\theta, \sigma / y) = \frac{L(\theta, \sigma) p(\theta, \sigma)}{\iint_{\theta \in \Theta} L(\theta, \sigma) p(\theta, \sigma) d\theta d\sigma} \quad (4.8)$$

onde Θ é o espaço paramétrico.

A função de verossimilhança $L(\theta, \sigma^2)$, para o modelo na forma matricial (4.5) é dada por,

$$\begin{aligned} L(\theta, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} [y - f(\theta)]' [y - f(\theta)]\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} S(\theta)\right). \end{aligned} \quad (4.9)$$

Combinando a informação dos dados (4.9) com a informação prévia (4.6), obtemos a densidade a posteriori conjunta para (θ, σ) dada por,

$$p(\theta, \sigma / y) \propto \sigma^{-(n+1)} \exp\left(-\frac{1}{2\sigma^2} S(\theta)\right). \quad (4.10)$$

Integrando (4.10) com respeito a σ e utilizando o resultado da distribuição gama invertida (Apêndice E (E.1)) temos a densidade a posteriori para θ , dada por,

$$p(\theta / y) \propto S(\theta)^{-n/2}, \quad (4.11)$$

ou

$$p(\theta / y) = k^{-1} S(\theta)^{-n/2}, \quad \theta \in \Theta.$$

onde,

$$k = \int_{\Theta} S(\theta)^{-n/2} d\theta \quad (4.12)$$

que representa a constante normalizadora é calculada através do uso de integração numérica. O resultado acima também é válido para o caso discreto substituindo-se a integral em (4.12) por uma soma.

Se (4.7) é substituída por $p(\theta, \sigma) \propto \sigma^{-1} p(\theta)$, temos,

$$p(\theta / y) \propto p(\theta) S(\theta)^{-n/2}, \quad (4.13)$$

que reduz-se a (4.11) quando $p(\theta)$ é constante.

A distribuição a posteriori (4.13) pode ser usada para estimação pontual (por exemplo, a média a posteriori $\mathcal{E}(\theta/y)$ pode ser usada como estimativa de θ) e para a construção de intervalos e regiões a posteriori (Katz et al.(1981)). Observamos ainda, que a moda de (4.13) é o valor de que minimiza $S(\theta)$, isto é, o estimador de mínimos quadrados (e de máxima verossimilhança) $\hat{\theta}$.

4.3.1. Momentos a Posteriori

Um estimador Bayesiano para o parâmetro de interesse é dado pela moda da densidade a posteriori. Uma outra forma de obter estimativas Bayesianas é considerar funções de perda para o parâmetro de interesse. Uma função de perda muito comum é a função de perda quadrática dada por,

$$\mathcal{L}(\theta) = (\theta - \hat{\theta})^2 \quad (4.14)$$

onde $\hat{\theta}$ é o estimador que minimiza a perda esperada com relação à densidade a posteriori. Considerando a função de perda quadrática, podemos verificar que o estimador Bayesiano para o parâmetro de interesse é dado pelo valor esperado da densidade a posteriori. Supor que desejamos obter momentos a posteriori para funções positivas $r(\theta)$ da forma:

$$\begin{aligned}
E(r) &= E[r(\theta) / y] \\
&= \int r(\theta) p(\theta / y) d\theta \\
&= k^{-1} \int r(\theta) p(y / \theta) p(\theta) d\theta \\
&= \frac{\int r(\theta) p(y / \theta) p(\theta) d\theta}{\int p(y / \theta) p(\theta) d\theta}
\end{aligned}$$

onde $r(\theta)$ é uma função seleccionada de θ , devemos ter em mente que uma das limitações dos métodos Bayesianos relaciona-se à resolução de integrais, que na maioria das vezes, não apresentam soluções analíticas, exigindo assim, o uso de métodos de aproximação (ver apêndice D).

Uma estimativa Bayesiana por intervalo pode ser obtida, através dos intervalos de credibilidade Bayesianos² aproximados, onde utilizamos métodos numéricos encontrando as áreas aproximadas abaixo da curva ($p(\theta/y)$), ou seja, determinamos numericamente os pontos (θ_*, θ^*) , tais que $P(\theta_* \leq \theta \leq \theta^*) = 1 - \alpha$. Assim, poderíamos ter,

$$\int_{-\infty}^{\theta_*} p(\theta / y) d\theta = \frac{\alpha}{2} ; \quad \int_{\theta^*}^{\infty} p(\theta / y) d\theta = \frac{\alpha}{2} \quad (4.15)$$

onde $p(\theta/y)$ é a densidade a posteriori unimodal de θ e (θ_*, θ^*) representa o intervalo Bayesiano que contém o verdadeiro valor de θ com probabilidade $(1-\alpha)$.

4.3.2. Densidades Preditivas

Em muitas ocasiões, dada a informação amostral, podemos ter interesse em fazer inferências sobre uma observação que ainda não foi observada, isto é, fazer uma previsão para uma observação futura. Na análise Bayesiana, a função densidade de probabilidade para uma observação ainda não observada, dada a informação amostral, pode ser obtida e é conhecida como densidade preditiva para uma observação futura, representada por y_{n+1}^* .

Sejam $Y_1, \dots, Y_n$ variáveis aleatórias com função densidade conhecida associada aos parâmetros $\theta_1, \dots, \theta_p$ que pertencem ao espaço paramétrico Θ . A densidade preditiva para uma observação futura y_{n+1}^* é dada por,

²Intervalos a posteriori baseados em prioris não-informativas são denominados "intervalos de credibilidade" por Edwards, Lindman e Savage (1963) e "intervalos de confiança Bayesianos" por Lindley (1965).

$$p(y_{n+1}^* / y_1, \dots, y_n) = \int_{\theta} p(y_{n+1}^* / \theta_1, \dots, \theta_p) p(\theta_1, \dots, \theta_p / y_1, \dots, y_n) d\theta_1 \dots d\theta_p \quad (4.16)$$

Dessa forma, $p(y_{n+1}^* / y_1, \dots, y_n) = E\{p(y_{n+1}^* / \theta_1, \dots, \theta_p)\}$, onde $p(y_{n+1}^* / \theta_1, \dots, \theta_p)$ é a função densidade de y_{n+1}^* e $p(\theta_1, \dots, \theta_p / y_1, \dots, y_n)$ é a densidade a posteriori dos parâmetros (para maiores informações ver, por exemplo, Aitchison e Dunsmore, 1975).

A seguir, apresentaremos os resultados da análise Bayesiana considerando a densidade a priori não-informativa de Jeffreys e a densidade a priori normal bivariada para o modelo de Michaelis-Menten.

4.4. Análise Bayesiana do Modelo de Michaelis-Menten considerando uma Densidade a Priori Não-Informativa

Nas seções seguintes, vamos obter inferências sobre os parâmetros do modelo através de uma análise Bayesiana para o modelo de Michaelis-Menten, considerando uma densidade a priori não-informativa (Box e Tiao, 1973), explorando métodos de aproximação de integrais, com especial destaque para o método de Laplace (ver por exemplo, Tierney e Kadane, 1986; ou Tierney, Kass e Kadane, 1989).

4.4.1. Formulação do Modelo

Considere o modelo não-linear de Michaelis-Menten dado por,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i \quad (4.17)$$

onde $y_i, i = 1, 2, \dots, n$ é a variável dependente, x_i é a variável independente, θ_1 e θ_2 são parâmetros e os erros ε_i são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal $N(0, \sigma^2)$.

Como vimos no capítulo anterior, uma aplicação especial desse modelo é dado em estudos de cinética de enzimas, onde y_i refere-se a velocidade inicial de uma reação

enzimática e x_i é a concentração de um substrato. Desejamos estimar $\theta_1 = V_{\max}$ a velocidade máxima e $\theta_2 = k_m$ a constante de Michaelis.

4.4.2. Uma Densidade a Priori Não-Informativa para θ e σ

A partir do momento em que observamos os dados, toda a informação sobre os parâmetros está contida na função de verossimilhança.

A função de verossimilhança para os parâmetros θ_1, θ_2 e σ^2 do modelo de Michaelis-Menten (4.17), considerando uma amostra aleatória de tamanho n é dada por,

$$L(\theta_1, \theta_2, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right\} \quad (4.18)$$

Uma densidade a priori para $\theta = (\theta_1, \theta_2)$ e σ , assumindo θ independente de σ (ver Seber e Wild, 1989) é dada por,

$$p(\theta_1, \theta_2, \sigma) = p(\theta_1, \theta_2) p_0(\sigma) \quad (4.19)$$

Assumindo que o pesquisador não tenha informação prévia a respeito dos parâmetros, vamos utilizar a regra multiparamétrica de Jeffreys (ver por exemplo, Box e Tiao, 1973), para a obtenção de uma priori não-informativa para θ_1 e θ_2 e considerar uma densidade a priori não-informativa para σ proporcional a σ^{-1} . Assim, uma densidade a priori conjunta para os parâmetros (ver (4.6)) é dada por,

$$p(\theta_1, \theta_2, \sigma) \propto \sigma^{-1} \left\{ \det(F'(\theta)F(\theta)) \right\}^{1/2}, \quad 0 < \theta_1, \theta_2 < \infty; \sigma > 0 \quad (4.20)$$

onde $\det(F'(\theta)F(\theta)) = \theta_1^2 g(\theta_2)$, e

$$g(\theta_2) = \left[\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} \right] \left[\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2.$$

Portanto, para o modelo de Michaelis-Menten (4.17), a densidade a priori de Jeffreys para θ_1, θ_2 e σ é dada por,

$$p(\theta_1, \theta_2, \sigma) \propto \theta_1 \sigma^{-1} [g(\theta_2)]^{1/2}, \quad 0 < \theta_1, \theta_2 < \infty; \sigma > 0 \quad (4.21)$$

A densidade a priori (4.21) coincide com a densidade a priori de Bernardo para os parâmetros (ver Eaves, 1983).

4.4.3. Densidade a Posteriori Conjunta para θ e σ

A densidade a posteriori conjunta para θ_1, θ_2 e σ , considerando a priori (4.21) e representando o vetor y (ver (4.8)) por "*dados*", é dada por,

$$p(\theta_1, \theta_2, \sigma / \text{dados}) \propto p(\theta_1, \theta_2, \sigma) L(\theta_1, \theta_2, \sigma^2) \quad (4.22)$$

onde $L(\theta_1, \theta_2, \sigma^2)$ é a função de verossimilhança (4.18) e $p(\theta_1, \theta_2, \sigma)$ representa a densidade a priori (4.21). A densidade a posteriori conjunta para os parâmetros do modelo (4.17), substituindo (4.21) e (4.18) em (4.22) é dada por,

$$p(\theta_1, \theta_2, \sigma / \text{dados}) \propto \theta_1 \sigma^{-(n+1)} [g(\theta_2)]^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right\} \quad (4.23)$$

Para integrar (4.23) em relação a σ , aplicamos o resultado (E.1) (ver apêndice E) e obtemos a densidade a posteriori conjunta para θ_1 e θ_2 , dada por,

$$p(\theta_1, \theta_2 / \text{dados}) \propto \theta_1 [g(\theta_2)]^{1/2} \left\{ \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right) \right\}^{-n/2}$$

Isto é,

$$p(\theta_1, \theta_2 / \text{dados}) \propto \theta_1 [g(\theta_2)]^{1/2} [S(\theta_1, \theta_2)]^{-n/2} \quad (4.24)$$

onde $0 < \theta_1, \theta_2 < \infty$ e $g(\theta_2)$ é definida na seção (4.4.2).

4.4.3.1. Densidades a Posteriori Marginais para θ_1 e θ_2

Nesta seção, vamos determinar as densidades a posteriori marginais para os parâmetros θ_1 e θ_2 do modelo de Michaelis-Menten (4.17).

Densidade a Posteriori Marginal para θ_1

A densidade a posteriori marginal para θ_1 é obtida integrando-se a densidade a posteriori conjunta (4.24) com respeito ao parâmetro θ_2 , isto é,

$$\begin{aligned} p(\theta_1 / \text{dados}) &\propto \int_0^{\infty} p(\theta_1, \theta_2 / \text{dados}) d\theta_2 \\ &\propto \theta_1 \int_0^{\infty} [g(\theta_2)]^{1/2} [S(\theta_1, \theta_2)]^{-n/2} d\theta_2 \end{aligned} \quad (4.25)$$

Como temos dificuldade na resolução analítica da integral em (4.25), vamos utilizar o método de Laplace de aproximação de integrais (ver apêndice D) e reescrever a densidade (4.25) na forma,

$$p(\theta_1 / \text{dados}) \propto \theta_1 \int e^{-nh_{\theta_1}(\theta_2)} d\theta_2 \quad (4.26)$$

onde $-nh_{\theta_1}(\theta_2) = \frac{1}{2} \ln g(\theta_2) - \frac{n}{2} \ln [S(\theta_1, \theta_2)]$.

Usando o método de Laplace para aproximação de integrais com $f(\theta_1, \theta_2) = 1$ na fórmula (D.2) do apêndice D, temos,

$$p(\theta_1 / \text{dados}) \propto \frac{\theta_1 [g(\hat{\theta}_2)]^{1/2} [S(\theta_1, \hat{\theta}_2)]^{-n/2}}{\left\{ \frac{nS_{22}(\theta_1, \hat{\theta}_2)}{S(\theta_1, \hat{\theta}_2)} - \frac{g''(\hat{\theta}_2)}{g(\hat{\theta}_2)} + \left(\frac{n-1}{n} \right) \cdot \left[\frac{g'(\hat{\theta}_2)}{g(\hat{\theta}_2)} \right]^2 \right\}^{1/2}} \quad (4.27)$$

onde $\theta_1 > 0, \hat{\theta}_2$ o máximo de $-nh_{\theta_1}(\theta_2)$ para cada valor fixo de θ_1 , é dado pela equação:

$$\frac{g'(\hat{\theta}_2)}{g(\hat{\theta}_2)} = \frac{nS_2(\theta_1, \hat{\theta}_2)}{S(\theta_1, \hat{\theta}_2)},$$

onde $S_i(\theta_1, \theta_2) = \partial S / \partial \theta_i, i = 1, 2; S_{ij}(\theta_1, \theta_2) = \partial^2 S / \partial \theta_i \partial \theta_j, i, j = 1, 2$.

Poderíamos considerar outras escolhas para $f(\theta)$ na fórmula (D.2) do apêndice D, com a finalidade de obter expressões mais simples e melhores precisões para as densidades a

posteriori marginais de interesse. Um caso especial, considerando a densidade a posteriori marginal para θ_1 é dado por,

$$p(\theta_1 / \text{dados}) \propto \theta_1 \int f(\theta_2) e^{-nh_{\theta_1}(\theta_2)} d\theta_2 \quad (4.28)$$

onde $f(\theta_2) = [g(\theta_2)]^{1/2}$ e $-nh_{\theta_1}(\theta_2) = -\frac{n}{2} \ln[S(\theta_1, \theta_2)]$.

Com esta escolha de $f(\theta)$, uma aproximação para a densidade a posteriori marginal de θ_1 é dada por,

$$p(\theta_1 / \text{dados}) \propto \frac{\theta_1^{1/2} [g(\hat{\theta}_2)]^{1/2} [S(\theta_1, \hat{\theta}_2)]^{\frac{-(n-1)}{2}}}{\left\{ 3\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^4} - 2 \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^3} \right\}^{1/2}} \quad (4.29)$$

onde $\theta_1 > 0$ e $\hat{\theta}_2$ é o máximo de $-nh_{\theta_1}(\theta_2) = -\frac{n}{2} \ln S(\theta_1, \theta_2)$ para cada valor fixado de θ_1 .

Densidade a Posteriori Marginal para θ_2

A densidade a posteriori marginal para θ_2 é dada pela integração da densidade a posteriori conjunta (4.24) em relação ao parâmetro θ_1 ,

$$p(\theta_2 / \text{dados}) \propto [g(\theta_2)]^{1/2} \int_0^{\infty} \theta_1 [S(\theta_1, \theta_2)]^{-n/2} d\theta_1 \quad (4.30)$$

Usando o método de Laplace com $f(\theta_1, \theta_2) = 1$ (ver apêndice D), podemos reescrever (4.30) na forma,

$$p(\theta_2 / \text{dados}) \propto [g(\theta_2)]^{1/2} \int e^{-nh_{\theta_2}(\theta_1)} d\theta_1 \quad (4.31)$$

onde $-nh_{\theta_2}(\theta_1) = \ln(\theta_1) - \frac{n}{2} \ln[S(\theta_1, \theta_2)]$.

Uma aproximação de Laplace para a densidade a posteriori marginal³ (4.31) é dada por,

³Para maiores detalhes sobre as expressões relacionadas com as densidades a posteriori marginais de θ_1 e θ_2 ver apêndice (E).

$$p(\theta_2 / \text{dados}) \propto \frac{\hat{\theta}_1 [g(\theta_2)]^{1/2} [S(\hat{\theta}_1, \theta_2)]^{-n/2}}{\left\{ \frac{nS_{11}(\hat{\theta}_1, \theta_2)}{2S(\hat{\theta}_1, \theta_2)} + \left(\frac{n-2}{n} \right) \frac{1}{\hat{\theta}_1^2} \right\}^{1/2}} \quad (4.32)$$

onde $\theta_2 > 0, \hat{\theta}_1$ o máximo de $-nh_{\theta_2}(\theta_1)$ para cada valor fixo de θ_2 é dado por,

$$\frac{2}{n\hat{\theta}_1} = \frac{S_1(\hat{\theta}_1, \theta_2)}{S(\hat{\theta}_1, \theta_2)}$$

4.4.4. Momentos a Posteriori Marginais para θ_1 e θ_2

Considerando a densidade a posteriori conjunta (4.24) para θ_1 e θ_2 , podemos utilizar o método de Laplace para aproximar momentos a posteriori para θ_1, θ_2 ou funções de θ_1 e θ_2 .

A média a posteriori para θ_1 é dada por,

$$E(\theta_1 / \text{dados}) = \frac{\iint_{\Theta} \theta_1 p(\theta_1, \theta_2 / \text{dados}) d\theta_1 d\theta_2}{\iint_{\Theta} p(\theta_1, \theta_2 / \text{dados}) d\theta_1 d\theta_2} = \frac{\iint_{\Theta} \theta_1^2 [g(\theta_2)]^{1/2} [S(\theta_1, \theta_2)]^{-n/2} d\theta_1 d\theta_2}{\iint_{\Theta} \theta_1 [g(\theta_2)]^{1/2} [S(\theta_1, \theta_2)]^{-n/2} d\theta_1 d\theta_2} \quad (4.33)$$

onde $\Theta = \{(\theta_1, \theta_2); 0 < \theta_1, \theta_2 < \infty\}$. Utilizando o método de Laplace reescrevemos (4.33) na forma,

$$E(\theta_1 / \text{dados}) = \frac{\iint e^{-nh^*(\theta_1, \theta_2)} d\theta_1 d\theta_2}{\iint e^{-nh(\theta_1, \theta_2)} d\theta_1 d\theta_2} \quad (4.34)$$

onde: $-nh^*(\theta_1, \theta_2) = 2 \ln(\theta_1) + \frac{1}{2} \ln g(\theta_2) - \frac{n}{2} \ln S(\theta_1, \theta_2)$ e

$$-nh(\theta_1, \theta_2) = \ln(\theta_1) + \frac{1}{2} \ln g(\theta_2) - \frac{n}{2} \ln S(\theta_1, \theta_2).$$

Assim, uma aproximação da média a posteriori para θ_1 é dada por,

$$\hat{E}(\theta_1 / \text{dados}) \cong \frac{\hat{\theta}_1^* [g(\hat{\theta}_2^*)]^{1/2} [S(\hat{\theta}_1^*, \hat{\theta}_2^*)]^{-n/2} \{\det nD^2 h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)\}^{-1/2}}{\hat{\theta}_1 [g(\hat{\theta}_2)]^{1/2} [S(\hat{\theta}_1, \hat{\theta}_2)]^{-n/2} \{\det nD^2 h(\hat{\theta}_1, \hat{\theta}_2)\}^{-1/2}} \quad (4.35)$$

onde $(\hat{\theta}_1^*, \hat{\theta}_2^*)$ maximizam $-nh^*(\theta_1, \theta_2)$; $(\hat{\theta}_1, \hat{\theta}_2)$ maximizam $-nh(\theta_1, \theta_2)$; $D^2h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)$ é a matriz Hessiana de h^* calculada em $(\hat{\theta}_1^*, \hat{\theta}_2^*)$ e $D^2h(\hat{\theta}_1, \hat{\theta}_2)$ é a matriz Hessiana de h calculada em $(\hat{\theta}_1, \hat{\theta}_2)$, (ver apêndice D).

Da mesma forma, podemos encontrar uma aproximação para a média a posteriori de θ_2 , dada pela expressão,

$$\hat{E}(\theta_2 / \text{dados}) \cong \frac{\hat{\theta}_1^* \hat{\theta}_2^* [g(\hat{\theta}_2^*)]^{1/2} [S(\hat{\theta}_1^*, \hat{\theta}_2^*)]^{-n/2} \{\det n D^2 h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)\}^{-1/2}}{\hat{\theta}_1 [g(\hat{\theta}_2)]^{1/2} [S(\hat{\theta}_1, \hat{\theta}_2)]^{-n/2} \{\det n D^2 h(\hat{\theta}_1, \hat{\theta}_2)\}^{-1/2}} \quad (4.36)$$

onde $(\hat{\theta}_1^*, \hat{\theta}_2^*)$ maximizam $-nh^*(\theta_1, \theta_2) = \ln(\theta_1) + \ln(\theta_2) + \frac{1}{2} \ln g(\theta_2) - \frac{n}{2} \ln S(\theta_1, \theta_2)$; $(\hat{\theta}_1, \hat{\theta}_2)$ maximizam $-nh(\theta_1, \theta_2) = \ln(\theta_1) + \frac{1}{2} \ln g(\theta_2) - \frac{n}{2} \ln S(\theta_1, \theta_2)$; $D^2h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)$ é a matriz Hessiana de h^* calculada em $(\hat{\theta}_1^*, \hat{\theta}_2^*)$ e $D^2h(\hat{\theta}_1, \hat{\theta}_2)$ é a matriz Hessiana de h calculada em $(\hat{\theta}_1, \hat{\theta}_2)$.¹

4.4.5. Densidade Preditiva para uma Observação Futura

A densidade preditiva de uma observação futura y_{n+1}^* , para um valor fixo x^* da variável independente, no modelo (4.17) dado os dados observados $(y_1, \dots, y_n)$, é definida por,

$$p(y_{n+1}^* / \text{dados}) = \iiint f(y_{n+1}^* / \theta_1, \theta_2, \sigma) p(\theta_1, \theta_2, \sigma / \text{dados}) d\theta_1 d\theta_2 d\sigma \quad (4.37)$$

onde $f(y_{n+1}^* / \theta_1, \theta_2, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*}\right)^2\right\}$ e $p(\theta_1, \theta_2, \sigma / \text{dados})$ representa a densidade a posteriori conjunta (4.23). Assim, a densidade preditiva para y_{n+1}^* é dada por,

$$p(y_{n+1}^* / \text{dados}) \propto \iiint_{\Theta} \theta_1 \sigma^{-(n+2)} [g(\theta_2)]^{1/2} \times \\ \times \exp\left\{-\frac{1}{2\sigma^2} \left[S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*}\right)^2\right]\right\} d\theta_1 d\theta_2 d\sigma$$

onde $\Theta = \{(\theta, \sigma); 0 < \theta_1, \theta_2 < \infty; \sigma > 0\}$.

¹Maiores informações sobre as expressões relacionadas com momentos a posteriori de θ_1 e θ_2 ver apêndice (E).

Integrando em relação a σ , (ver resultado D.1 do apêndice D), obtemos,

$$p(y_{n+1}^* / \text{dados}) \propto \iint_{\Theta} \frac{\theta_1 [g(\theta_2)]^{1/2} d\theta_1 d\theta_2}{\left\{ S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*} \right)^2 \right\}^{\frac{(n+1)}{2}}} \quad (4.38)$$

onde $\Theta = \{(\theta_1, \theta_2); 0 < \theta_1, \theta_2 < \infty\}$. Usando o método de Laplace com $f = [g(\theta_2)]^{1/2}$, (ver (D.2) no apêndice D), obtemos uma aproximação para a densidade preditiva para y_{n+1}^* dada por,

$$p(y_{n+1}^* / \text{dados}) \propto \frac{\hat{\theta}_1 [g(\hat{\theta}_2)]^{1/2} \left\{ \det n D^2 h(\hat{\theta}_1, \hat{\theta}_2) \right\}^{-1/2}}{\left\{ S(\hat{\theta}_1, \hat{\theta}_2) + \left(y_{n+1}^* - \frac{\hat{\theta}_1 x^*}{\hat{\theta}_2 + x^*} \right)^2 \right\}^{\frac{(n+1)}{2}}} \quad (4.39)$$

onde $(\hat{\theta}_1, \hat{\theta}_2)$ maximizam $-nh(\theta_1, \theta_2) = \ln(\theta_1) - \left(\frac{n+1}{2} \right) \ln \left\{ S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*} \right)^2 \right\}$ para cada valor fixado de y_{n+1}^* e $D^2 h(\hat{\theta}_1, \hat{\theta}_2)$ é a matriz Hessiana de h calculada em $(\hat{\theta}_1, \hat{\theta}_2)$.²

4.5. Análise Bayesiana do Modelo de Michaelis-Menten considerando uma densidade a Priori Informativa

Nesta seção, apresentaremos uma análise Bayesiana do modelo de Michaelis-Menten, considerando densidades a priori informativas para representar o grau de conhecimento sobre os parâmetros θ_1 e θ_2 .

4.5.1. Uma Densidade a Priori Informativa para θ_1, θ_2 e σ

A informação de estudos preliminares ou do próprio pesquisador pode ser utilizada como uma forma para melhorar os resultados inferenciais, isto é, utilizando informação a priori.

²Maiores informações sobre as expressões utilizadas na seção (4.4.5) ver apêndice (E).

Utilizando o modelo de Michaelis-Menten, vamos considerar densidades a priori informativas para representar o grau de conhecimento sobre os parâmetros θ_1, θ_2 e σ . A priori conjunta para θ_1, θ_2 e σ pode ser escrita por,

$$p(\theta_1, \theta_2, \sigma / a, b, \sigma_1, \sigma_2, \rho) = p(\theta_1, \theta_2 / a, b, \sigma_1, \sigma_2, \rho) \cdot p_0(\sigma) \quad (4.40)$$

onde $\sigma > 0$; $0 < \theta_1, \theta_2 < \infty$; $a, b, \sigma_1, \sigma_2, \rho$ são constantes conhecidas..

Consideramos uma densidade a priori para θ_1, θ_2 e σ dada (de (4.40)) pelo produto de uma distribuição normal bivariada para (θ_1, θ_2) com médias e variâncias conhecidas e uma distribuição a priori não-informativa para σ . Assim,

$$p(\theta_1, \theta_2 / a, b, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{NORM\}$$

com

$$NORM = -\frac{1}{2(1-\rho^2)} [(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]$$

$$N1 = \frac{\theta_1 - a}{\sigma_1}$$

e

$$N2 = \frac{\theta_2 - b}{\sigma_2}$$

para $0 < \theta_1, \theta_2 < \infty$, e

$$p_0(\sigma) \propto \frac{1}{\sigma}$$

para $\sigma > 0$.

Assim, a densidade a priori conjunta para θ_1, θ_2 e σ é dada de (4.40) por,

$$p(\theta_1, \theta_2, \sigma / a, b, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{NORM\} \quad (4.41)$$

onde $0 < \theta_1, \theta_2 < \infty$ e $\sigma > 0$.

4.5.2. Densidade a Posteriori Conjunta para θ_1, θ_2 e σ

Considerando a funo de verossimilhana dada em (4.18) e a priori informativa (4.41), a densidade a posteriori conjunta para θ_1, θ_2 e σ (ver por exemplo, Box e Tiao, 1973)  dada por,

$$p(\theta_1, \theta_2, \sigma / \text{dados}) \propto \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]\right\} \times \\ \times \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right\} \quad (4.42)$$

onde $\sigma > 0$; $0 < \theta_1, \theta_2 < \infty$.

4.5.3. Densidades a Posteriori Marginais para θ_1, θ_2

A densidade a posteriori conjunta para (θ_1, θ_2)  obtida integrando-se a densidade (4.42) em relao a σ . Assim, a densidade a posteriori conjunta para (θ_1, θ_2)  dada por,

$$p(\theta_1, \theta_2 / \text{dados}) \propto \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]\right\} \times \\ \times [S(\theta_1, \theta_2)]^{-\frac{n}{2}} \quad (4.43)$$

onde $0 < \theta_1, \theta_2 < \infty$ e $S(\theta_1, \theta_2) = \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2$.

As densidades a posteriori marginais para θ_1 e θ_2 so obtidas pela integrao da densidade a posteriori conjunta (4.43) em relao a θ_2 e θ_1 , respectivamente. Assim, a densidade a posteriori marginal para θ_1 , considerando a densidade a priori informativa (4.41), pode ser expressa na forma,

$$p(\theta_1 / \text{dados}) \propto \int_0^\infty f(\theta_1, \theta_2) e^{-nh(\theta_1, \theta_2)} d\theta_2 \quad (4.44)$$

onde:

$$f(\theta_1, \theta_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]\right\}$$

e

$$-nh(\theta_1, \theta_2) = -\frac{n}{2} \ln[S(\theta_1, \theta_2)].$$

Como a integral em (4.44) não apresenta solução analítica explícita, utilizamos a aproximação de Laplace (ver Apêndice D) e encontramos a densidade a posteriori marginal aproximada para θ_1 , dada por,

$$p(\theta_1 / \text{dados}) \propto \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]\right\} \times \\ \times \left\{ \frac{n}{2} \left[\frac{S_{22}(\theta_1, \hat{\theta}_2)}{S(\theta_1, \hat{\theta}_2)} - \left(\frac{S_2(\theta_1, \hat{\theta}_2)}{S(\theta_1, \hat{\theta}_2)} \right)^2 \right] \right\}^{-\frac{1}{2}} \times \{S(\theta_1, \hat{\theta}_2)\}^{-\frac{n}{2}} \quad (4.45)$$

onde $0 < \theta_1 < \infty$;

$$S_{22}(\theta_1, \hat{\theta}_2) = -4\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^3} + 6\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^4};$$

$$S_2(\theta_1, \hat{\theta}_2) = 2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\hat{\theta}_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\hat{\theta}_2 + x_i)^3}$$

e $\hat{\theta}_2$ maximiza $-nh(\theta_1, \theta_2)$ e é dado pela solução da equação,

$$\frac{S_2(\theta_1, \hat{\theta}_2)}{S(\theta_1, \hat{\theta}_2)} = 0$$

A densidade a posteriori marginal para θ_2 , considerando a densidade a priori informativa (4.41), é dada por,

$$p(\theta_2 / \text{dados}) \propto \int_0^{\infty} f(\theta_1, \theta_2) e^{-nh(\theta_1, \theta_2)} d\theta_1 \quad (4.46)$$

onde $f(\theta_1, \theta_2)$ e $-nh(\theta_1, \theta_2)$ estão definidas em (4.44).

Similarmente, utilizamos a aproximação de Laplace em (4.46) e encontramos a densidade a posteriori marginal para θ_2 dada por,

$$p(\theta_2 / \text{dados}) \propto \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2]\right\} \times \\ \times \left\{\frac{n}{2}\left[\frac{S_{11}(\hat{\theta}_1, \theta_2)}{S(\hat{\theta}_1, \theta_2)} - \left(\frac{S_1(\hat{\theta}_1, \theta_2)}{S(\hat{\theta}_1, \theta_2)}\right)^2\right]\right\}^{-\frac{1}{2}} \times \{S(\hat{\theta}_1, \theta_2)\}^{-\frac{n}{2}} \quad (4.47)$$

onde $0 < \theta_1 < \infty$;

$$S_{11}(\hat{\theta}_1, \theta_2) = 2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2};$$

$$S_1(\hat{\theta}_1, \theta_2) = -2 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} + 2\hat{\theta}_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2}$$

e $\hat{\theta}_1$ maximiza $-nh(\theta_1, \theta_2)$ e é dado pela solução da equação,

$$\frac{S_1(\hat{\theta}_1, \theta_2)}{S(\hat{\theta}_1, \theta_2)} = 0$$

4.5.4. Densidade Preditiva para uma Observação Futura

A densidade preditiva para uma observação futura y_{n+1}^* (ver por exemplo, Press, 1989), considerando densidades a priori informativas é dada por,

$$p(y_{n+1}^* / \text{dados}) = \iiint_{\Theta} f(y_{n+1}^* / \theta_1, \theta_2, \sigma) p(\theta_1, \theta_2, \sigma / \text{dados}) d\theta_1 d\theta_2 d\sigma$$

onde $\Theta = \{0 < \theta_1, \theta_2, \sigma < \infty\}$; $f(y_{n+1}^* / \theta_1, \theta_2, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*}\right)^2\right\}$ e

$p(\theta_1, \theta_2, \sigma / \text{dados})$ representa a densidade a posteriori conjunta (4.42). Assim, a densidade preditiva para y_{n+1}^* é dada por,

$$p(y_{n+1}^* / \text{dados}) \propto \iiint_{\Theta} \sigma^{-(n+2)} \exp \left\{ -\frac{1}{2\sigma^2} \left[S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*} \right)^2 \right] \right\} \times \\ \times \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2] \right\} d\theta_1 d\theta_2 d\sigma$$

onde $\Theta = \{(\theta_1, \theta_2, \sigma); 0 < \theta_1, \theta_2, \sigma < \infty\}$.

Integrando a densidade acima em relação a σ , (ver resultado D.1 do Apêndice D), obtemos,

$$p(y_{n+1}^* / \text{dados}) \propto \iint_{\Theta} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2] \right\} \times \\ \times \left\{ S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*} \right)^2 \right\}^{-\frac{(n+1)}{2}} d\theta_1 d\theta_2 \quad (4.48)$$

onde $\Theta = \{(\theta_1, \theta_2); 0 < \theta_1, \theta_2 < \infty\}$.

Utilizando o método de aproximação de integrais de Laplace, com

$$f(\theta_1, \theta_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [(N1)^2 - 2\rho(N1) \cdot (N2) + (N2)^2] \right\}$$

obtemos uma aproximação para a densidade preditiva para y_{n+1}^* dada por,

$$p(y_{n+1}^* / \text{dados}) \propto \left\{ \det nD^2 h(\hat{\theta}_1, \hat{\theta}_2) \right\}^{-\frac{1}{2}} \cdot \left\{ S(\hat{\theta}_1, \hat{\theta}_2) + \left(y_{n+1}^* - \frac{\hat{\theta}_1 x^*}{\hat{\theta}_2 + x^*} \right)^2 \right\}^{-\frac{(n+1)}{2}} \cdot f(\hat{\theta}_1, \hat{\theta}_2) \quad (4.49)$$

onde $(\hat{\theta}_1, \hat{\theta}_2)$ maximizam $-nh(\theta_1, \theta_2) = -\frac{(n+1)}{2} \ln \left\{ S(\theta_1, \theta_2) + \left(y_{n+1}^* - \frac{\theta_1 x^*}{\theta_2 + x^*} \right)^2 \right\}$ para cada valor fixo de y_{n+1}^* e $D^2 h(\hat{\theta}_1, \hat{\theta}_2)$ é a matriz Hessiana de h calculada em $(\hat{\theta}_1, \hat{\theta}_2)$.

A expressão (4.49) pode ser utilizada para obter estimativas pontuais ou estimativas por intervalos da previsão y_{n+1}^* ; fixando o valor desejado da concentração de substrato em x^* .

4.6. Reparametrizaes para o Modelo de Michaelis-Menten

No contexto Bayesiano  importante verificar se a utilizao de diferentes reparametrizaes proporciona melhores aproximaes para as densidades a posteriori de interesse ou alguma vantagem computacional.

Para isso, consideramos as reparametrizaes propostas por Bates e Watts (1988) dada em (3.1) e a proposta por Seber e Wild (1989) dada em (3.2). Tambm consideramos a priori de Jeffreys (4.21) para θ_1, θ_2 e σ .

Inicialmente, vamos considerar a reparametrizao proposta por Bates e Watts (1988), dada por,

$$\begin{cases} \beta_1 = \frac{\theta_2}{\theta_1} \\ \beta_2 = \frac{1}{\theta_1} \end{cases}, \quad \text{isto ,} \quad \begin{cases} \theta_1 = \frac{1}{\beta_2} \\ \theta_2 = \frac{\beta_1}{\beta_2} \end{cases} \quad (4.50)$$

A matriz Jacobiana dessa transformao  dada por,

$$\det\{J(\beta_1, \beta_2)\} = \begin{vmatrix} \frac{\partial \theta_1}{\partial \beta_1} & \frac{\partial \theta_1}{\partial \beta_2} \\ \frac{\partial \theta_2}{\partial \beta_1} & \frac{\partial \theta_2}{\partial \beta_2} \end{vmatrix} = \begin{vmatrix} 0 & -1 \\ \frac{1}{\beta_2} & \frac{-\beta_1}{\beta_2^2} \end{vmatrix}$$

Portanto, o $\det\{J(\beta_1, \beta_2)\} = \frac{1}{\beta_2^3}$.

Usando a tcnica de transformao de variveis, a densidade a posteriori conjunta para β_1 e β_2  (de (4.24)) dada por,

$$p(\beta_1, \beta_2 / \text{dados}) \propto |J| p(\theta_1, \theta_2 / \text{dados})$$

Assim,

$$p(\beta_1, \beta_2 / \text{dados}) \propto \frac{1}{\beta_2^4} [g(\beta_1, \beta_2)]^{1/2} \left\{ \sum_{i=1}^n \left(y_i - \frac{x_i}{\beta_1 + \beta_2 x_i} \right)^2 \right\}^{-n/2} \quad (4.51)$$

onde $0 < \beta_1 < \infty$; $0 < \beta_2 < \infty$ e

$$g(\beta_1, \beta_2) = \left[\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^2} \right] \cdot \left[\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^4} \right] - \left[\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^3} \right]^2.$$

Integrando (4.51) em relação a β_2 , obtemos a densidade a posteriori marginal de β_1 , dada por,

$$p(\beta_1 / \text{dados}) \propto \int_0^\infty \frac{1}{\beta_2^4} [g(\beta_1, \beta_2)]^{1/2} \left\{ \sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \beta_2 x_i)} \right)^2 \right\}^{-n/2} d\beta_2 \quad (4.52)$$

Utilizando o método de Laplace de aproximação de integrais (ver apêndice D, para maiores detalhes), temos,

$$p(\beta_1 / \text{dados}) \propto \int e^{-nh_{\beta_1}(\beta_2)} d\beta_2$$

$$\text{onde } -nh_{\beta_1}(\beta_2) = -4 \ln(\beta_2) + \frac{1}{2} \ln[g(\beta_1, \beta_2)] - \frac{n}{2} \ln \left\{ \sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \beta_2 x_i)} \right)^2 \right\}.$$

Assim, a densidade a posteriori marginal aproximada para β_1 , obtida pelo método de Laplace, é dada por,

$$p(\beta_1 / \text{dados}) \propto \frac{1}{\hat{\beta}_2^4} [g(\beta_1, \hat{\beta}_2)]^{1/2} \left\{ \sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \hat{\beta}_2 x_i)} \right)^2 \right\}^{-n/2} \{A(\beta_1, \hat{\beta}_2)\}^{-1/2} \quad (4.53)$$

onde $0 < \beta_1 < \infty$;

$$\begin{aligned}
A(\beta_1, \hat{\beta}_2) = & -\frac{4}{\hat{\beta}_2^2} - \frac{g''(\beta_1, \hat{\beta}_2)}{2g(\beta_1, \hat{\beta}_2)} + \frac{1}{2} \left(\frac{g'(\beta_1, \hat{\beta}_2)}{g(\beta_1, \hat{\beta}_2)} \right)^2 + \\
& + \left\{ \frac{-2n \sum_{i=1}^n \frac{x_i^3 y_i}{(\beta_1 + \hat{\beta}_2 x_i)^3} + 3n \sum_{i=1}^n \left(\frac{x_i^4}{(\beta_1 + \hat{\beta}_2 x_i)^4} \right)}{\left[\sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \hat{\beta}_2 x_i)} \right)^2 \right]} \right\} - \\
& - \left\{ \frac{\left[n \sum_{i=1}^n \frac{x_i^2 y_i}{(\beta_1 + \hat{\beta}_2 x_i)^2} - n \sum_{i=1}^n \frac{x_i^3}{(\beta_1 + \hat{\beta}_2 x_i)^3} \right] \cdot \left[2 \sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \hat{\beta}_2 x_i)} \right) \cdot \left(\frac{x_i^2}{(\beta_1 + \hat{\beta}_2 x_i)^2} \right) \right]}{\left[\sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \hat{\beta}_2 x_i)} \right)^2 \right]^2} \right\}
\end{aligned}$$

e $\hat{\beta}_2$ maximiza $-nh_{\beta_1}(\beta_2)$ dado em (4.53).

Similarmente, obtemos a densidade a posteriori marginal para β_2 , dada por,

$$p(\beta_2 / \text{dados}) \propto \frac{1}{\beta_2^4} [g(\hat{\beta}_1, \beta_2)]^{1/2} \left\{ \sum_{i=1}^n \left(y_i - \frac{x_i}{(\hat{\beta}_1 + \beta_2 x_i)} \right)^2 \right\}^{-n/2} \{B(\hat{\beta}_1, \beta_2)\}^{-1/2} \quad (4.54)$$

onde $0 < \beta_2 < \infty$; $\hat{\beta}_1$ maximiza $-nh_{\beta_2}(\beta_1) = \frac{1}{2} \ln g(\beta_1, \beta_2) - \frac{n}{2} \ln \sum_{i=1}^n \left(y_i - \frac{x_i}{(\beta_1 + \beta_2 x_i)} \right)^2$ e

$$\begin{aligned}
B(\hat{\beta}_1, \beta_2) = & -\frac{g''(\hat{\beta}_1, \beta_2)}{g(\hat{\beta}_1, \beta_2)} + \frac{1}{2} \left(\frac{g'(\hat{\beta}_1, \beta_2)}{g(\hat{\beta}_1, \beta_2)} \right)^2 + \\
& + \left\{ \frac{-2n \sum_{i=1}^n \frac{x_i y_i}{(\hat{\beta}_1 + \beta_2 x_i)^3} + 3n \sum_{i=1}^n \frac{x_i^2}{(\hat{\beta}_1 + \beta_2 x_i)^4}}{\sum_{i=1}^n \left(y_i - \frac{x_i}{(\hat{\beta}_1 + \beta_2 x_i)} \right)^2} \right\} - \left\{ \frac{2 \left[\sum_{i=1}^n \frac{x_i y_i}{(\hat{\beta}_1 + \beta_2 x_i)^2} - \sum_{i=1}^n \frac{x_i^2}{(\hat{\beta}_1 + \beta_2 x_i)^3} \right]^2}{\left[\sum_{i=1}^n \left(y_i - \frac{x_i}{(\hat{\beta}_1 + \beta_2 x_i)} \right)^2 \right]^2} \right\}
\end{aligned}$$

A verificação da melhoria da normalidade das densidades a posteriori de interesse e da função de verossimilhança, para esta reparametrização, está feita com os dados no Capítulo 7.

Considerando a reparametrização proposta por Seber e Wild (1989), dada em (3.2), a matriz Jacobiana da transformação é dada por,

$$J(\delta_1, \delta_2) = \begin{vmatrix} \frac{\delta_2(u-v)}{(u\delta_2 - v\delta_1)} + \frac{\delta_1\delta_2 v(u-v)}{(u\delta_2 - v\delta_1)^2} & \frac{\delta_1(u-v)}{(u\delta_2 - v\delta_1)} - \frac{\delta_1\delta_2 u(u-v)}{(u\delta_2 - v\delta_1)^2} \\ \frac{uv}{(u\delta_2 - v\delta_1)} + \frac{uv^2(\delta_1 - \delta_2)}{(u\delta_2 - v\delta_1)^2} & \frac{uv}{(u\delta_2 - v\delta_1)} - \frac{u^2 v(\delta_1 - \delta_2)}{(u\delta_2 - v\delta_1)^2} \end{vmatrix}$$

Portanto,

$$\det\{J(\delta_1, \delta_2)\} = \frac{uv\delta_1\delta_2(u-v) \cdot (u+v)}{(u\delta_2 - v\delta_1)^3} + \frac{uv(u-v) \cdot (\delta_2 - \delta_1)}{(u\delta_2 - v\delta_1)^2} - \frac{uv(\delta_1 - \delta_2) \cdot (u-v) \cdot (u\delta_2 + v\delta_1)}{(u\delta_2 - v\delta_1)^3}$$

Desta forma, a densidade a posteriori conjunta para (δ_1, δ_2) é dada por,

$$p(\delta_1, \delta_2 / \text{dados}) \propto \det\{J(\delta_1, \delta_2)\} \frac{\delta_1\delta_2(u-v)}{(u\delta_2 - v\delta_1)} \sqrt{g(\delta_1, \delta_2)} \times \left\{ \sum_{i=1}^n \left(y_i - \frac{(u-v)x_i\delta_1\delta_2}{v(u-x_i)\delta_1 - u(v-x_i)\delta_2} \right)^2 \right\}^{-\frac{n}{2}} \quad (4.55)$$

onde $0 < \delta_1, \delta_2 < \infty$ e

$$g(\delta_1, \delta_2) = \left\{ \sum_{i=1}^n \frac{x_i^2 (u\delta_2 - v\delta_1)^2}{[uv(\delta_1 - \delta_2) + x_i(u\delta_2 - v\delta_1)]^2} \right\} \cdot \left\{ \sum_{i=1}^n \frac{x_i^2 (u\delta_2 - v\delta_1)^4}{[uv(\delta_1 - \delta_2) + x_i(u\delta_2 - v\delta_1)]^4} \right\} - \left\{ \sum_{i=1}^n \frac{x_i^2 (u\delta_2 - v\delta_1)^3}{[uv(\delta_1 - \delta_2) + x_i(u\delta_2 - v\delta_1)]^3} \right\}^2$$

Para calcular as densidades a posteriori marginais para δ_1 e δ_2 , devemos utilizar métodos numéricos ou de aproximações de integrais.

A densidade a posteriori marginal para δ_1 é obtida na forma,

$$p(\delta_1 / \text{dados}) \propto \int_0^\infty \det\{J(\delta_1, \delta_2)\} \frac{\delta_1 \delta_2 (u-v)}{(u\delta_2 - v\delta_1)} \sqrt{g(\delta_1, \delta_2)} \times \\ \times \left\{ \sum_{i=1}^n \left(y_i - \frac{(u-v)x_i \delta_1 \delta_2}{v(u-x_i)\delta_1 - u(v-x_i)\delta_2} \right)^2 \right\}^{-\frac{n}{2}} d\delta_2 \quad (4.56)$$

De forma análoga, obtemos a densidade a posteriori marginal para δ_2 , dada por,

$$p(\delta_2 / \text{dados}) \propto \int_0^\infty \det\{J(\delta_1, \delta_2)\} \frac{\delta_1 \delta_2 (u-v)}{(u\delta_2 - v\delta_1)} \sqrt{g(\delta_1, \delta_2)} \times \\ \times \left\{ \sum_{i=1}^n \left(y_i - \frac{(u-v)x_i \delta_1 \delta_2}{v(u-x_i)\delta_1 - u(v-x_i)\delta_2} \right)^2 \right\}^{-\frac{n}{2}} d\delta_1 \quad (4.57)$$

Como as integrais em (4.56) e (4.57) não apresentam soluções analíticas explícitas, para resolvê-las, usamos métodos numéricos ou de aproximações. Uma forma simples de calcular essas densidades a posteriori é através do método numérico de Simpson (ver Apêndice D). Também devemos recorrer a métodos numéricos para a obtenção das esperanças a posteriori de δ_1 e δ_2 , visando a simplificação dos cálculos.

Com esta reparametrização proposta por Seber e Wild, (1989), verificamos uma melhoria significativa na normalidade da função de verossimilhança e nas densidades a posteriori de interesse. A verificação está feita com os dados no Capítulo 7.

Capítulo 5

O MODELO DE MICHAELIS-MENTEN SEM A SUPOSIÇÃO DE NORMALIDADE PARA O ERRO

Nos capítulos anteriores, apresentamos as teorias clássica e Bayesiana para os modelos não-lineares onde consideramos os erros com distribuição normal em problemas convencionais. Embora as interpretações sejam diferentes, tanto os resultados Bayesianos como os resultados da teoria clássica são precisos. É importante salientar, que qualquer um dos métodos de inferência discutidos requerem certas suposições necessárias para garantir resultados precisos, especificamente, as suposições de independência e normalidade para as observações.

Como a inferência Bayesiana não é restrita no que diz respeito às limitações técnicas da teoria amostral, o objetivo principal deste capítulo é a investigação de um modo no qual esta flexibilidade possa ser utilizada. Em particular, vamos estudar as consequências do relaxamento das suposições convencionais para os modelos não-lineares.

Neste capítulo, vamos supor que os erros não sejam normalmente distribuídos. Especificamente, vamos considerar problemas de inferência sobre os coeficientes de regressão do modelo não-linear de Michaelis-Menten, através de uma classe mais ampla de distribuições para os erros que inclui a distribuição Normal como um caso especial.

5.1. A escolha da Classe de Distribuições Potência Exponencial

A análise de dados com modelos lineares ou não-lineares, frequentemente, é feita sob a suposição de normalidade para os erros ε_i . Na análise Bayesiana existe a possibilidade de estudar o efeito de não-normalidade nas inferências, estudando os efeitos de suposições distribucionais mais amplas. Portanto, vamos considerar uma distribuição mais ampla para os erros, uma distribuição que seja membro da classe de distribuições simétricas que inclui a distribuição Normal, como um caso especial, juntamente com outras distribuições leptocúrticas e platicúrticas, para estudar o efeito da não-normalidade nas inferências. A distribuição normal padronizada pode ser escrita na forma,

$$p(x) = k \exp\left(-\frac{1}{2}|x|^q\right) \quad \text{onde } q=2.$$

Assumindo outros valores para q , obtemos uma classe de distribuições mais ampla, a classe de distribuições potência exponencial. Estas distribuições são consideradas por Diananda (1949), Box (1953b) e Turner (1960); com $q = \frac{2}{1+\beta}$ temos uma densidade dada na forma,

$$\begin{aligned} p(y/\theta, \beta, \phi) &= k\phi^{-1} \exp\left(-\frac{1}{2}\left|\frac{y-\theta}{\phi}\right|^{2/(1+\beta)}\right) \\ &= k\phi^{-1} \exp\left\{-\frac{1}{2}\left[\left(\frac{y-\theta}{\phi}\right)^2\right]^{1/(1+\beta)}\right\} \end{aligned} \quad (5.1)$$

onde

$$k^{-1} = \Gamma\left(1 + \frac{1+\beta}{2}\right) \cdot 2^{1+\frac{1}{2}(1+\beta)} \text{ e } \phi > 0; -\infty < \theta < \infty; -\infty < y < \infty \text{ e } -1 < \beta \leq 1.$$

Na expressão (5.1), θ é o parâmetro de locação e ϕ o parâmetro de escala. Podemos, facilmente, mostrar que,

$$E(y) = \theta$$

e

$$Var(y) = \sigma^2 = 2^{(1+\beta)} \left\{ \frac{\Gamma\left[\frac{3}{2}(1+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]} \right\} \phi^2$$

Alternativamente, podemos expressar (5.1) na forma,

$$p(y/\theta, \sigma, \beta) = w(\beta)\sigma^{-1} \exp\left[-c(\beta) \cdot \left|\frac{y-\theta}{\sigma}\right|^{\frac{2}{1+\beta}}\right], \quad -\infty < y < \infty \quad (5.2)$$

onde:

$$c(\beta) = \left\{ \frac{\Gamma\left(\frac{3}{2}(1+\beta)\right)}{\Gamma\left(\frac{1}{2}(1+\beta)\right)} \right\}^{1/(1+\beta)}$$

e

$$w(\beta) = \frac{\left\{ \Gamma\left[\frac{3}{2}(1+\beta)\right] \right\}^{\frac{1}{2}}}{(1+\beta) \left\{ \Gamma\left[\frac{1}{2}(1+\beta)\right] \right\}^{\frac{3}{2}}}, \quad \sigma > 0, -\infty < \theta < \infty, -1 < \beta \leq 1.$$

Os parâmetros θ e σ são, respectivamente, a média e o desvio padrão da distribuição. O parâmetro β é uma medida de curtose, um indicador de não-normalidade. Em particular, quando $\beta = 0$, temos uma distribuição normal e quando $\beta = 1$, temos uma distribuição exponencial dupla com densidade,

$$p(y / \theta, \sigma, \beta = 1) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2} \left| \frac{y - \theta}{\sigma} \right| \right), \quad -\infty < y < \infty$$

Quando β tende a -1 , temos uma distribuição uniforme, isto é,

$$\lim_{\beta \rightarrow -1} p(y / \beta, \theta, \sigma) = \frac{1}{2\sqrt{3}\sigma}, \quad \theta - \sqrt{3} \cdot \sigma < y < \theta + \sqrt{3} \cdot \sigma$$

Na figura (5.1), observamos a forma da distribuição potência exponencial para vários valores de β . As distribuições apresentadas têm média e desvio padrão comuns. Observamos graficamente, que para $\beta > 0$ as distribuições são leptocúrticas e para $\beta < 0$ as distribuições são platicúrticas.

Portanto, vamos supor para o modelo de Michaelis-Menten (4.17), que os erros sejam variáveis aleatórias independentes com distribuição potência exponencial dada por,

$$p(\varepsilon / \sigma, \beta) = w(\beta) \sigma^{-1} \exp\left\{ -c(\beta) \left| \frac{\varepsilon}{\sigma} \right|^{2/(1+\beta)} \right\} \quad (5.3)$$

onde $-\infty < \varepsilon < \infty$; $\sigma > 0$; $-1 < \beta \leq 1$; $w(\beta)$ e $c(\beta)$ estão definidos em (5.2).

Assim, vamos explorar o uso desta classe ampla de distribuições simétricas (5.3) para os erros ε_i , onde o parâmetro β , que é uma medida de curtose (indicador de não-normalidade), assume o valor zero para a distribuição normal. A opinião sobre a distribuição dos dados pode ser expressa atribuindo a β uma densidade a priori informativa unimodal centrada em zero. A acurácia dessa densidade a priori pode ser alterada, tal que o efeito da variação do grau de incerteza sobre a normalidade possa ser representado.

Esta é uma forma na qual a pesquisa Bayesiana pode ser uma ferramenta útil e muito poderosa para avaliar o efeito de incertezas nas suposições.

Figura 5.1. Distribuições potência exponencial para vários valores de β .

5.2. Critério de Robustez e Inferência Robusta

Supondo que as observações $y_1, y_2, \dots, y_n$ representem uma amostra aleatória de uma população normal $N(\theta, \sigma^2)$, as inferências sobre θ são baseadas no uso de critérios da teoria normal. Em circunstâncias práticas, é usual justificar o uso desses critérios, nas quais a normalidade não pode ser garantida, argumentando-se que a distribuição de critérios da teoria normal é pouco afetada pela não-normalidade da distribuição dos dados, ou seja, é robusto sob não-normalidade. Assim, temos um critério de robustez, sob não-normalidade.

Embora o critério da teoria normal possa não ser muito alterado sem a suposição de normalidade para os dados, a inferência a ser deduzida de uma dada amostra pode ser muito

diferente quando utilizamos um critério apropriado para uma outra distribuição. Para distinguir do critério de robustez, à propriedade de insensibilidade das inferências, denominamos de inferência robusta.

Neste capítulo, vamos verificar a robustez ou não das inferências, considerando o modelo de Michaelis-Menten (4.17) com a densidade (5.3) para o erro. Para diferentes valores de β , verificamos se as curvas ajustadas diferem entre si, o que indica uma sensibilidade das inferências obtidas, considerando inicialmente, β conhecido. No caso em que β é suposto desconhecido, vamos verificar se existem diferenças significativas nas distribuições a posteriori de θ_1 e θ_2 .

5.3. Análise Bayesiana do Modelo de Michaelis-Menten considerando β conhecido

Assumindo uma distribuição potência exponencial (5.3) para os erros ε_i do modelo de Michaelis-Menten (4.17) e uma densidade a priori não-informativa (ver Box e Tiao, 1973) para os parâmetros do modelo, desenvolvemos, neste contexto, uma análise Bayesiana usando o método de Laplace para aproximar integrais (ver por exemplo, Tierney e Kadane, 1986) quando não conseguimos achar soluções analíticas explícitas para as densidades a posteriori marginais de interesse. Inicialmente, consideramos que o parâmetro indicador de não-normalidade β seja conhecido.

5.3.1. Densidade a Priori Não-Informativa para θ_1, θ_2 e σ

Dado uma amostra de n observações independentes de um membro da classe de distribuições (5.3), a função de verossimilhança é dada por,

$$L(\theta_1, \theta_2, \sigma, \beta / \text{dados}) \propto [w(\beta)]^n \sigma^{-n} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{\varepsilon_i}{\sigma} \right|^{\frac{2}{(1+\beta)}} \right] \quad (5.4)$$

onde $\varepsilon_i = y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}$.

Assumindo β conhecido, podemos expressar a função de verossimilhança (5.4), na forma,

$$L(\theta_1, \theta_2, \sigma / \beta, \text{dados}) \propto \sigma^{-n} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{\varepsilon_i}{\sigma} \right|^{\frac{2}{1+\beta}} \right] \quad (5.5)$$

Para representar a opinião sobre os parâmetros θ_1 e θ_2 , assumimos (θ_1, θ_2) e $\log(\sigma)$ aproximadamente independentes e localmente uniformes a priori (ver por exemplo, Box e Tiao, 1973), isto é, com uma densidade a priori não-informativa,

$$p(\theta_1, \theta_2, \sigma) \propto \frac{1}{\sigma} \quad (5.6)$$

$\sigma > 0$.

5.3.2. Densidade a Posteriori Conjunta para θ_1, θ_2 e σ

Considerando a densidade a priori não-informativa (5.6) e a função de verossimilhança (5.5), a densidade a posteriori conjunta para θ_1, θ_2 e σ é dada por, (ver Box e Tiao, 1973),

$$p(\theta_1, \theta_2, \sigma / \beta, \text{dados}) \propto \sigma^{-(n+1)} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{\varepsilon_i}{\sigma} \right|^{\frac{2}{1+\beta}} \right] \quad (5.7)$$

onde $0 < \theta_1, \theta_2 < \infty; \sigma > 0$.

5.3.3. Densidade a Posteriori Conjunta para θ_1 e θ_2

A densidade a posteriori conjunta para (θ_1, θ_2) é obtida integrando-se a densidade a posteriori conjunta (5.7) com respeito a σ . Assim, obtemos,

$$p(\theta_1, \theta_2 / \beta, \text{dados}) = [J(\beta)]^{-1} [M(\theta_1, \theta_2)]^{-\frac{1}{2}n(1+\beta)} \quad (5.8)$$

onde $0 < \theta_1, \theta_2 < \infty$, $M(\theta_1, \theta_2) = \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{(1+\beta)}}$ e a constante normalizadora é definida por,

$$J(\beta) = \iint_{\Theta} [M(\theta_1, \theta_2)]^{-\frac{n}{2}(1+\beta)} d\theta_1 d\theta_2$$

onde $\Theta = \{0 < \theta_j < \infty, j = 1, 2\}$.

Assim, para cada valor fixo de β , vamos considerar, a densidade a posteriori conjunta para (θ_1, θ_2) dada por,

$$p(\theta_1, \theta_2 / \beta, \text{dados}) \propto [M(\theta_1, \theta_2)]^{-\frac{n}{2}(1+\beta)} \quad (5.9)$$

onde $0 < \theta_1, \theta_2 < \infty$.

Na prática, devemos calcular a moda $(\hat{\theta}_1, \hat{\theta}_2)$ de (5.9) para diferentes valores de β ($-1 < \beta \leq 1$) para verificar a robustez ou não das inferências obtidas.

5.4. Análise Bayesiana do Modelo de Michaelis-Menten considerando β desconhecido

Devido às diferenças significativas que possam ocorrer nas distribuições a posteriori de θ_1 e θ_2 , dependendo da distribuição dos dados utilizada, surge uma questão: qual seria o melhor valor para β no supermodelo considerado? Existem duas formas para a obtenção de informações sobre o valor de β , baseando-se na informação dos dados ou através do conhecimento a priori.

Podemos assim, inserir cada informação na análise Bayesiana considerando β como uma variável aleatória associada com uma distribuição a priori $p(\beta)$.

5.4.1. Densidade a Priori Não-Informativa para $\theta_1, \theta_2, \sigma$ e β

Assumimos que, a priori, β seja independente de $(\theta_1, \theta_2$ e $\sigma)$, (ver Box e Tiao, 1973), tal que,

$$p(\theta_1, \theta_2, \sigma, \beta) = p(\theta_1, \theta_2, \sigma) p(\beta)$$

Considerando para θ_1, θ_2 e σ , uma densidade a priori não-informativa dada por,

$$p(\theta_1, \theta_2, \sigma) \propto \frac{1}{\sigma}$$

temos,

$$p(\theta_1, \theta_2, \sigma, \beta) \propto \sigma^{-1} p(\beta) \quad (5.10)$$

Uma densidade a priori de referência para β pode ser a distribuição uniforme definida em $(-1, 1)$. Assim, uma densidade a priori para os parâmetros $\theta_1, \theta_2, \sigma$ e β de (5.10), é dada por,

$$p(\theta_1, \theta_2, \sigma, \beta) \propto \frac{1}{\sigma} \quad (5.11)$$

onde $\sigma > 0$.

5.4.2. Densidade a Posteriori Conjunta para $\theta_1, \theta_2, \sigma$ e β

Considerando a densidade a priori (5.11) e a função de verossimilhança (5.5), a densidade a posteriori conjunta para $\theta_1, \theta_2, \sigma$ e β (ver por exemplo, Box e Tiao, 1973), é dada por,

$$p(\theta_1, \theta_2, \sigma, \beta / \text{dados}) \propto [w(\beta)]^n \sigma^{-(n+1)} \exp \left\{ -c(\beta) \sigma^{-2/(1+\beta)} M(\theta_1, \theta_2) \right\} \quad (5.12)$$

$$\text{onde } M(\theta_1, \theta_2) = \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{1/(1+\beta)} ; 0 < \theta_1, \theta_2 < \infty ; \sigma > 0 ; -1 < \beta \leq 1.$$

5.4.3. Densidade a Posteriori Conjunta para θ_1, θ_2 , e β

A densidade a posteriori conjunta para θ_1, θ_2 e β é obtida integrando-se a densidade a posteriori conjunta (5.12) com respeito a σ , isto é,

$$p(\theta_1, \theta_2, \beta / \text{dados}) \propto [w(\beta)]^n \int_0^\infty \sigma^{-(n+1)} \exp\{-a\sigma^{-\alpha}\} d\sigma \quad (5.13)$$

onde:

$$\alpha = \frac{2}{1+\beta}; \quad a = c(\beta)M(\theta_1, \theta_2).$$

Usando o resultado (E.1 do Apêndice E), a densidade a posteriori conjunta para θ_1, θ_2 e β é dada por,

$$p(\theta_1, \theta_2, \beta / \text{dados}) \propto [w(\beta)]^n (1+\beta) [c(\beta)]^{-\frac{n}{2}(1+\beta)} \times \\ \times [M(\theta_1, \theta_2)]^{-\frac{n}{2}(1+\beta)} \Gamma\left[\frac{n}{2}(1+\beta)\right] \quad (5.14)$$

onde $0 < \theta_1, \theta_2 < \infty$; $-1 < \beta \leq 1$; $w(\beta)$ e $c(\beta)$ estão definidos em (5.2).

5.4.4. Densidade a Posteriori Marginal para β

A densidade a posteriori marginal para β , que é obtida integrando-se a densidade a posteriori conjunta (5.14) com respeito a θ_1 e θ_2 , é dada por,

$$p_u(\beta / \text{dados}) \propto [w(\beta)]^n (1+\beta) [c(\beta)]^{-\frac{n}{2}(1+\beta)} \Gamma\left[\frac{n}{2}(1+\beta)\right] \times \\ \times \iint_{\Theta} [M(\theta_1, \theta_2)]^{-\frac{n}{2}(1+\beta)} d\theta_1 d\theta_2 \quad (5.15)$$

onde $\Theta = \{0 < \theta_j < \infty, j = 1, 2\}$.

Como não temos uma solução analítica explícita para a integral em (5.15), usamos o método de Laplace para aproximação de integrais (ver por exemplo, Tierney e Kadane,

(1986); ou Kass, Tierney e Kadane, (1990)), para calcular a densidade a posteriori marginal para β .

Assim, escrevemos,

$$p_u(\beta / \text{dados}) \propto [w(\beta)]^n (1+\beta) [c(\beta)]^{-\frac{n}{2}(1+\beta)} \Gamma\left[\frac{n}{2}(1+\beta)\right] \times \int_{\Theta} e^{-nh(\theta_1, \theta_2)} d\theta_1 d\theta_2 \quad (5.16)$$

onde $\Theta = \{0 < \theta_j < \infty, j = 1, 2\}$ e $-nh(\theta_1, \theta_2) = -\frac{n}{2}(1+\beta) \ln[M(\theta_1, \theta_2)]$.

A densidade a posteriori marginal aproximada para β é dada por,

$$p_u(\beta / \text{dados}) \propto [w(\beta)]^n (1+\beta) [c(\beta)]^{-\frac{n}{2}(1+\beta)} \Gamma\left[\frac{n}{2}(1+\beta)\right] \times \frac{[M(\hat{\theta}_1, \hat{\theta}_2)]^{-\frac{n}{2}(1+\beta)+1}}{\left\{M_{11}(\hat{\theta}_1, \hat{\theta}_2)M_{22}(\hat{\theta}_1, \hat{\theta}_2) - [M_{12}(\hat{\theta}_1, \hat{\theta}_2)]^2\right\}^{\frac{1}{2}}} \quad (5.17)$$

onde $-1 < \beta \leq 1$; $(\hat{\theta}_1, \hat{\theta}_2)$ maximizam $-nh(\theta_1, \theta_2)$ para cada valor fixo de β e são dados a partir das equações,

$$\begin{cases} \frac{M_1(\hat{\theta}_1, \hat{\theta}_2)}{M(\hat{\theta}_1, \hat{\theta}_2)} = 0 \\ \frac{M_2(\hat{\theta}_1, \hat{\theta}_2)}{M(\hat{\theta}_1, \hat{\theta}_2)} = 0 \end{cases}$$

onde

$$M_1(\theta_1, \theta_2) = \frac{1}{(1+\beta)} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{(1+\beta)}-1} \cdot a_{1i}$$

$$M_2(\theta_1, \theta_2) = \frac{1}{(1+\beta)} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-1} \cdot a_{2i}$$

$$M_{11}(\theta_1, \theta_2) = -\frac{\beta}{(1+\beta)^2} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-2} \cdot a_{1i}^2 + \\ + \frac{1}{(1+\beta)} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-1} \cdot a_{11i}$$

$$M_{22}(\theta_1, \theta_2) = -\frac{\beta}{(1+\beta)^2} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-2} \cdot a_{2i}^2 + \\ + \frac{1}{(1+\beta)} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-1} \cdot a_{22i}$$

$$M_{12}(\theta_1, \theta_2) = -\frac{\beta}{(1+\beta)^2} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-2} \cdot a_{1i} \cdot a_{2i} + \\ + \frac{1}{(1+\beta)} \sum_{i=1}^n \left[\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]^{\frac{1}{1+\beta}-1} \cdot a_{12i}$$

e

$$a_{1i} = -\frac{2x_i}{(\theta_2 + x_i)} \cdot \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)$$

$$a_{2i} = \frac{2\theta_1 x_i}{(\theta_2 + x_i)^2} \cdot \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)$$

$$a_{11i} = \frac{2x_i^2}{(\theta_2 + x_i)^2}$$

$$a_{22i} = \frac{2\theta_1^2 x_i^2}{(\theta_2 + x_i)^4} - \frac{4\theta_1 x_i}{(\theta_2 + x_i)^3} \cdot \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)$$

$$a_{12i} = -\frac{2\theta_1 x_i^2}{(\theta_2 + x_i)^3} + \frac{2x_i}{(\theta_2 + x_i)^2} \cdot \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right).$$

Supondo que o pesquisador deseje associar uma grande probabilidade a priori para distribuições dos dados próximas da distribuição normal, podemos, então, escolher uma distribuição a priori para β concentrada em $\beta = 0$. Box e Tiao (1973) consideram, a fim de representar este propósito, uma priori informativa Beta simétrica, para o parâmetro de interesse β , com média zero e variação de -1 a $+1$, com um parâmetro ajustável a . Especificamente, assumimos que,

$$p(\beta) \propto (1 - \beta^2)^{a-1}, \quad -1 < \beta \leq 1 \quad (5.18)$$

Quando $a = 1$, (5.18) é uma distribuição uniforme. Para $a > 1$, (5.18) é uma distribuição simétrica tendo sua moda, o valor $\beta = 0$ e tornando-se mais concentrada em $\beta = 0$ quando a aumenta. Para $a \rightarrow \infty$, $p(\beta)$ se aproxima de uma função delta, representando uma suposição de normalidade exata (ver Box e Tiao, 1973).

Assim, a densidade a posteriori marginal para β , considerando a densidade a priori informativa (5.18) e a densidade (5.17), é dada por,

$$p(\beta / \text{dados}) \propto p(\beta) p_u(\beta / \text{dados}) \quad (5.19)$$

onde $-1 < \beta \leq 1$.

Para um valor específico de a , toda a inferência sobre (θ_1, θ_2) , pode ser feita através das distribuições condicionais $p(\theta_j / \beta, \text{dados})$ com uma densidade, $p(\beta / \text{dados})$, apropriada, onde:

$$p(\theta_j / \text{dados}) \propto \int p(\theta_1, \theta_2 / \beta, \text{dados}) p_u(\beta / \text{dados}) p(\beta) d\beta$$

$0 < \theta_j < \infty; j = 1, 2$.

Neste caso, onde consideramos β desconhecido, várias situações podem ocorrer:

i) $p(\theta / \beta, \text{dados})$ pode ser insensível a β , assim, não é necessário o conhecimento exato do parâmetro *nuisance* β .

ii) $p(\theta / \beta, \text{dados})$ pode ser sensível a β e a integração desta densidade condicional com a função peso $p(\beta / \text{dados}) \propto p(\beta) p_u(\beta / \text{dados})$ deve ser estudada detalhadamente.

iii) Se $p_u(\beta / \text{dados})$ for precisa (quando o número de observações for grande), as inferências deverão apoiar-se na informação amostral e pouco nas escolhas de $p(\beta)$. Caso contrário, as inferências dependem praticamente das escolhas de $p(\beta)$.

Capítulo 6

TRANSFORMAÇÕES DOS DADOS

Em regressão linear bem como em regressão não-linear, frequentemente assumimos que as observações $y_1, y_2, \dots, y_n$ são independentes, normalmente distribuídas com variância constante e média especificada pelo modelo com um conjunto de parâmetros θ .

Embora usualmente assumimos que,

$$y_i = f(x_i, \theta) + \varepsilon_i \quad (i = 1, 2, \dots, n)$$

onde os ε_i são i.i.d. $N(0, \sigma^2)$, frequentemente existem violações da estrutura estocástica do modelo, particularmente com respeito à distribuição e homogeneidade da variância dos erros ε_i . Se um modelo é ajustado, por exemplo, por mínimos quadrados, os resíduos

$$\begin{aligned} \hat{\varepsilon}_i &= y_i - \hat{y}_i \\ &= y_i - f(x_i, \hat{\theta}) \end{aligned}$$

podem indicar a invalidade do modelo e a invalidade das suposições sobre os erros. Os resíduos podem ser ilusórios, particularmente se o modelo é altamente não-linear (como indicado pela curvatura intrínseca, ver seção (1.7.2.2.2)). Estes resíduos são chamados de resíduos "projetados" e têm propriedades similares aos resíduos ordinários para modelos lineares e podem ser utilizados para diagnosticar a não-normalidade e heterogeneidade da variância. Quando as suposições são claramente violadas, um modelo mais complexo deve ser construído. Este processo é contínuo até encontrarmos um modelo apropriado e este modelo final é então utilizado para fazer inferências sobre os dados.

Neste capítulo, vamos considerar a família paramétrica de transformações de potência (ver Box e Cox, 1964) com o objetivo de encontrar, após a aplicação de uma transformação adequada às observações y_i , um modelo apropriado para inferir sobre a transformação e sobre os parâmetros do modelo não-linear, sob o enfoque Bayesiano, através de densidades a posteriori relevantes.

6.1. Uma Família Paramétrica de Transformações de Potência

Um método geral, que fornece inferências estatísticas sobre a escolha de uma transformação, foi introduzido por Box e Cox (1964) que consideram a família paramétrica de transformações de potência definida da forma,

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \ln(y) & \text{se } \lambda = 0 \end{cases} \quad (6.1)$$

para $y > 0$. Observamos que, na ausência de uma transformação, $\lambda=1$. O valor da transformação para $\lambda = 0$ é o limite de (6.1) quando λ tende a zero. Utilizando a expansão de Taylor, temos as seguintes aproximações de primeira ordem,

$$\begin{aligned} E[y^{(\lambda)}] &\approx \{E[y]\}^{(\lambda)} \\ &\approx \{f(x_i, \theta)\}^{(\lambda)} \end{aligned}$$

e

$$\begin{aligned} Var[y^{(\lambda)}] &\approx Var[y] \left[\frac{dy^{(\lambda)}}{dy} \right]_{y=E(y)}^2 \\ &= Var[y] [E(y)]^{2\lambda-2} \end{aligned} \quad (6.2)$$

Assim, podemos escolher λ para tornar (6.2) constante.

Observamos que para modelos lineares, o principal interesse é encontrar λ em (6.1), tal que,

$$E[y^{(\lambda)}] = \mathbf{X}\theta$$

onde $y^{(\lambda)}$ satisfaz as suposições de variância constante, independência, aditividade e são normalmente distribuídos. Se a suposição de aditividade é encontrada, então a estrutura de $\mathbf{X}\theta$ será relativamente simples.

A transformação de potência (6.1) é somente uma das muitas famílias paramétricas de transformações que são sugeridas na literatura para a análise de dados. Neste capítulo,

desenvolvemos a teoria para transformação das variáveis respostas que são não-negativas, sob o enfoque Bayesiano para um modelo de Michaelis-Menten. Quando a variável resposta y assumir valor negativo, Box e Cox (1964) sugerem uma transformação alternativa.

6.2. Transformação da Variável Dependente

Quando não dispomos de normalidade para o erro no modelo de Michaelis-Menten,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i \quad (6.3)$$

podemos considerar a família de transformação de potência (ver Box e Cox, 1964) definida em (6.1) para a variável dependente y .

Em geral, assumimos que para cada λ , $y^{(\lambda)}$ é uma função monotônica de y sobre o espaço paramétrico admissível. Supor que observamos um vetor $(n \times 1)$ de observações $y' = (y_1, y_2, \dots, y_n)$ e que o modelo não-linear apropriado é especificado por,

$$E[y^{(\lambda)}] = \left(\frac{\theta_1 x_i}{\theta_2 + x_i} \right) \quad (6.4)$$

onde $y^{(\lambda)}$ é o vetor coluna de observações transformadas, x_i representa a variável independente e $\theta' = (\theta_1, \theta_2)$ é o vetor de parâmetros desconhecidos associados com as observações transformadas.

Assumimos que para um valor de λ desconhecido, as observações transformadas $y^{(\lambda)}$ satisfazem as suposições usuais, isto é, são independentes e normalmente distribuídas com variância constante σ^2 e com médias dadas por (6.4). A função densidade de probabilidade para as observações não-transformadas e consequentemente a função de verossimilhança em relação às observações originais é obtida multiplicando-se a densidade normal pelo jacobiano da transformação.

Portanto, a função de verossimilhança em relação às observações y é dada por,

$$L(\theta, \sigma^2, \lambda) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \prod_{i=1}^n \left\{ \exp\left(-\frac{1}{2\sigma^2} [y_i^{(\lambda)} - f(x_i, \theta)^{(\lambda)}]^2\right) \right\} \cdot J(\lambda, y) \quad (6.5)$$

onde $J(\lambda, y)$ é o jacobiano da transformação dado na forma,

$$J(\lambda, y) = \prod_{i=1}^n \left| \frac{dy_i^{(\lambda)}}{dy_i} \right|.$$

Assim, podemos inferir sobre os parâmetros em (6.5) utilizando a teoria de máxima verossimilhança para grandes amostras que fornece diretamente estimativas pontuais para os parâmetros, testes e intervalos de confiança aproximados, baseados na distribuição Qui-Quadrado (ver Box e Cox, 1964) ou através da análise Bayesiana.

6.3. Análise Bayesiana para o modelo de Michaelis-Menten

Sob o enfoque Bayesiano, podemos fazer inferências sobre os parâmetros em (6.5) utilizando as densidades a posteriori de interesse, para o modelo de Michaelis-Menten.

6.3.1. Densidade a Priori para λ , θ e σ

De (6.5), o logaritmo da função de verossimilhança para os parâmetros do modelo (6.3), excluindo as constantes, é dado por,

$$l(\theta_1, \theta_2, \sigma^2, \lambda) \propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} S^{(\lambda)}(\theta_1, \theta_2) + \ln J(\lambda, y)$$

$$\text{onde } S^{(\lambda)}(\theta_1, \theta_2) = \sum_{i=1}^n \left[y_i^{(\lambda)} - \left(\frac{\theta_1 x_i}{\theta_2 + x_i} \right)^{(\lambda)} \right]^2 \text{ e } J(\lambda, y) = \prod_{i=1}^n \left| \frac{dy_i^{(\lambda)}}{dy_i} \right| = \prod_{i=1}^n y_i^{\lambda-1}.$$

Assim, temos,

$$l(\theta_1, \theta_2, \sigma^2, \lambda) \propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} S^{(\lambda)}(\theta_1, \theta_2) + (\lambda - 1) \sum_{i=1}^n \ln(y_i) \quad (6.6)$$

Para uma análise Bayesiana, a densidade a priori para λ , θ e σ pode ser escrita na forma,

$$p(\lambda, \theta, \sigma) = p(\theta, \sigma / \lambda) p_o(\lambda) \quad (6.7)$$

Dado λ , vamos considerar a densidade a priori para θ_1, θ_2 e σ dada em (4.21). Portanto, assumindo uma densidade a priori uniforme para λ , uma densidade a priori conjunta para $\lambda, \theta_1, \theta_2$ e σ é dada (de (6.7)) por,

$$p(\lambda, \theta_1, \theta_2, \sigma) \propto \frac{\theta_1}{\sigma} [g(\theta_2)]^{\frac{1}{2}} \quad (6.8)$$

onde $g(\theta_2)$ está definida em (4.20).

6.3.2. Densidade a Posteriori Conjunta para λ, θ e σ

Considerando a densidade a priori não-informativa (6.8) e a função de verossimilhança (6.5), a densidade a posteriori conjunta para $\lambda, \theta_1, \theta_2$ e σ é dada (ver por exemplo, Box e Tiao, 1973) por,

$$p(\lambda, \theta_1, \theta_2, \sigma / \text{dados}) \propto \sigma^{-(n+1)} \theta_1 [g(\theta_2)]^{\frac{1}{2}} \left(\prod_{i=1}^n y_i^{\lambda-1} \right) \exp \left\{ -\frac{1}{2} S^{(\lambda)}(\theta_1, \theta_2) \right\} \quad (6.9)$$

onde $-\infty < \lambda < \infty$ e $\theta_1, \theta_2 > 0$.

Integrando (6.9) em relação a σ e utilizando o resultado (E.1.) do apêndice (E), obtemos a densidade a posteriori conjunta para θ_1, θ_2 e λ dada por,

$$p(\lambda, \theta_1, \theta_2 / \text{dados}) \propto \theta_1 [g(\theta_2)]^{\frac{1}{2}} \left(\prod_{i=1}^n y_i^{\lambda-1} \right) \cdot \{S^{(\lambda)}(\theta_1, \theta_2)\}^{-\frac{n}{2}} \quad (6.10)$$

onde $-\infty < \lambda < \infty$ e $\theta_1, \theta_2 > 0$.

6.3.3. Densidade a Posteriori Marginal para λ

A densidade a posteriori marginal de λ , é dada (de (6.10)) por,

$$p(\lambda / \text{dados}) \propto \left(\prod_{i=1}^n y_i^{\lambda-1} \right) \iint \theta_1 [g(\theta_2)]^{\frac{1}{2}} \{S^{(\lambda)}(\theta_1, \theta_2)\}^{-\frac{n}{2}} d\theta_1 d\theta_2 \quad (6.11)$$

onde $\theta_1, \theta_2 > 0$.

Como a integral em (6.11) não apresenta solução analítica explícita, usamos métodos numéricos ou de aproximação para resolvê-la. Uma forma simples de se obter a densidade a posteriori marginal para λ é através do método de Simpson (ver Apêndice D).

Um estimador pontual para λ , é dado pela moda $\hat{\lambda}$ da densidade a posteriori (6.11). Assim, com a transformação das variáveis dependentes $y^{(\hat{\lambda})}$, podemos utilizar a teoria proposta nos capítulos precedentes deste trabalho, para assim, obter as inferências para os parâmetros do modelo $y_i^{(\hat{\lambda})} = f(x_i, \theta) + \varepsilon_i$, onde os ε_i passam a ser i.i.d. $N(0, \sigma^2)$ e $f(x_i, \theta)$ representa o modelo de Michaelis-Menten (6.3), com a variável dependente y transformada.

Capítulo 7

APLICAÇÕES

7.1. Exemplo 1

Considerando o modelo de Michaelis-Menten, dado por:

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, \quad (7.1)$$

onde y_i , $i = 1, 2, \dots, n$ é a variável dependente, x_i é a variável independente ou regressora, θ_1 e θ_2 são os parâmetros do modelo e os erros ε_i são variáveis aleatórias i.i.d. com distribuição normal $N(0, \sigma^2)$. Uma aplicação especial do modelo (7.1) é dada no estudo de cinética de enzimas, onde y refere-se a velocidade inicial de uma reação enzimática e x representa a concentração de substrato. Neste caso, desejamos estimar $\theta_1 = V$ ou $V_{\max}$, a velocidade máxima e $\theta_2 = k_m$, a constante de Michaelis. Neste contexto, o modelo (7.1) é conhecido como equação de Michaelis-Menten (ver por exemplo, Raaijmakers, 1987).

Na tabela (1), temos os dados sobre a velocidade da reação enzimática, obtidos por Treloar (1974). O número de "counts" por minuto de produto radioativo da reação foi medido como uma função da concentração de substrato em partes por milhão (ppm) e destes "counts" foi calculada a razão inicial, ou velocidade da reação (counts/min²). O experimento foi conduzido com as enzimas tratadas com o substrato Puromycin. A velocidade é assumida como dependente da concentração de substrato de acordo com a equação de Michaelis-Menten.

Derivando y em relação aos parâmetros θ_1 e θ_2 do modelo (7.1), obtemos,

$$\begin{aligned} \frac{\partial y}{\partial \theta_1} &= \frac{x}{\theta_2 + x} \\ \frac{\partial y}{\partial \theta_2} &= \frac{-\theta_1 x}{(\theta_2 + x)^2} \end{aligned}$$

Como as derivadas envolvem pelo menos um dos parâmetros, o modelo (7.1) é reconhecido como não-linear.

<i>Concentração de Substrato</i> (ppm) x	<i>Velocidade da reação</i> (counts/min ²) y
0.02	76
0.02	47
0.06	97
0.06	107
0.11	123
0.11	139
0.22	159
0.22	152
0.56	191
0.56	201
1.10	207
1.10	200

Tabela 1. Dados da velocidade da reação versus a concentração de substrato para o experimento Puromycin.

Através da figura (1), observamos o gráfico das variáveis velocidade da reação versus a concentração de substrato.

Figura 1. Gráfico da velocidade da reação versus a concentração de substrato para os dados Puromycin.

Análise Clássica

Nesta seção, aplicamos aos dados da tabela (1), a metodologia de análise clássica baseada na aproximação normal assintótica dos estimadores de máxima verossimilhança e na análise clássica proposta nos capítulos (1 e 2).

Considerando os dados da tabela (1) e as expressões dadas em (2.30), os estimadores de máxima verossimilhança para θ_1 e θ_2 (obtidos pelo método iterativo de Newton-Raphson) e para σ^2 , são dados por: $\hat{\theta}_1 = 212.684$; $\hat{\theta}_2 = 0.06412$ e $\hat{\sigma}^2 = 99.6208$.

A matriz de variância e covariância assintótica de $\hat{\theta}_1, \hat{\theta}_2$ e $\hat{\sigma}^2$, a inversa da matriz de informação de Fisher (2.33), é dada por,

$$I^{-1}(\theta_1, \theta_2, \sigma^2) = \begin{pmatrix} 40.2190 & 0.0367 & 0 \\ 0.0367 & 0.0001 & 0 \\ 0 & 0 & 1654.005 \end{pmatrix}$$

Intervalos de confiança 95% aproximados para θ_1 e θ_2 , considerando a aproximação normal assintótica (2.31) dos e.m.v. $\hat{\theta}_1$ e $\hat{\theta}_2$, são dados por, (200.2539 ; 225.1140) e (0.0445; 0.0837), respectivamente.

Aplicando o método de Gauss-Newton (ver Capítulo 1, seção (1.4.3)) obtemos os estimadores de mínimos quadrados para os parâmetros θ_1 e θ_2 , com 7 iterações para a convergência. As estimativas são dadas por, $\hat{\theta}_1 = 212.684$, $E.P.(\hat{\theta}_1) = 6.94715$ e $\hat{\theta}_2 = 0.06412$, $E.P.(\hat{\theta}_2) = 0.00828$ e a soma de quadrados residual é dada por $SQR = 1195.45$. Assim, um estimador da variância é dado por,

$$s^2 = \frac{SQR(\hat{\theta})}{n-p} = \frac{1195.45}{10} = 119.545$$

Adequabilidade do modelo

Sabemos que a construção do modelo de regressão está baseada nas seguintes afirmações:

- os resíduos são independentes e identicamente distribuídos com distribuição normal;
- a variância dos resíduos é constante;
- a média dos resíduos é zero.

Para verificar se o modelo é adequado para os dados, ou seja, se as afirmações são válidas, um estudo detalhado sobre os resíduos e sobre o coeficiente de explicação R^2 é necessário.

Considerando o modelo de Michaelis-Menten (7.1) e os dados da tabela (1), o modelo ajustado pelo método de Gauss-Newton é dado por,

$$\hat{y}_i = \frac{212.684x_i}{0.06412 + x_i}, i = 1, 2, \dots, 12 \quad (7.2)$$

As suposições básicas são: $\varepsilon_i: i.i.d. N(0, \sigma^2)$.

Na figura (2), observando o gráfico dos resíduos versus os valores ajustados (a), notamos que os pontos estão distribuídos aleatoriamente em torno de zero, implicando que a variância dos resíduos é constante e a média é igual a zero. O gráfico normal probabilístico (b) apresenta boa linearidade, indicando, portanto, boa normalidade para os resíduos. Ressaltando que o número de observações ($n = 12$) é muito pequeno, notamos a presença de uma observação um pouco afastada, mas não é considerada como um outlier.

Figura 2. Gráfico dos resíduos do modelo (7.1) versus os valores ajustados e Gráfico Normal Probabilístico.

O coeficiente de explicação do modelo, $R^2 = 0.9613$, indica-nos que o modelo explica 96.1% da variabilidade dos dados, o que é muito satisfatório, e assim, podemos concluir que o modelo proposto (7.2) é adequado para os dados.

Como usualmente, as análises estatísticas para os modelos não-lineares são baseadas em resultados assintóticos, a precisão desses resultados pode ser verificada através de um

estudo de simulação (Ratkowsky, 1983), pelas medidas de não-linearidade, (Bates e Watts, 1980) e através da medida de vício proposta por Box (1971).

O estudo de simulação (ver apêndice C) pode ser útil para verificar o comportamento não-normal dos estimadores. Assim, vamos investigar o comportamento não-linear do modelo examinando as propriedades amostrais dos estimadores de mínimos quadrados. Foram simuladas 1000 amostras de tamanho 12 e para cada conjunto de dados, ajustamos o modelo por mínimos quadrados, usando o método numérico de Gauss-Newton. Assim, determinamos os momentos amostrais para $\hat{\theta}_1$ e $\hat{\theta}_2$.

<i>Momentos Amostrais para $\hat{\theta}_1$</i>			
<i>Momentos Amostrais</i>		<i>Quantidades Derivadas</i>	
Média	212.959	%Vício	0.1293%
Variância	45.785	% exc. variância	-0.089%
3.º Momento	10.853	Skewness	0.03503
4.º Momento	6671.780	exc. Curtose	0.18270

Tabela 2. Momentos amostrais para $\hat{\theta}_1$.

As quantidades derivadas de Skewness e o coeficiente de Curtose não são significativas ao nível de significância $\alpha = 0.05$. A porcentagem que excede a variância e a porcentagem do vício não são significativas, pois, como lei empírica, só consideramos vícios acima de 1% como sendo significativos.

Figura 3. Histograma para $\hat{\theta}_1$ e $\hat{\theta}_2$, respectivamente.

Observando o histograma na figura (3.a) e a tabela (2), concluímos que a distribuição de $\hat{\theta}_1$ é próxima da distribuição normal.

A tabela (3) contém os momentos amostrais de $\hat{\theta}_2$ e observando-a podemos concluir que a porcentagem do vício é igual a 1%, o que indica que o vício é insignificante e a porcentagem que excede a variância também não é significativa. Considerando $\alpha = 0.05$, podemos concluir que o coeficiente de Curtose não é significativo e o coeficiente de Skewness ultrapassa o valor crítico (ver apêndice C), mas esse excesso não é grande, pois através do histograma para $\hat{\theta}_2$, figura (3.b), observamos uma distribuição aproximadamente simétrica, de modo que podemos considerar o coeficiente de Skewness como não significativo.

<i>Momentos Amostrais para $\hat{\theta}_2$</i>			
<i>Momentos Amostrais</i>		<i>Quantidades Derivadas</i>	
Média	0.06483	% Vício	1%
Variância	0.0000697	% exc.variância	-0.0279%
3.ºMomento	0.000000125	Skewness	0.2148
4.ºMomento	0.000000016	exc. Curtose	0.2914

Tabela 3. Momentos Amostrais para $\hat{\theta}_2$.

Assim, da tabela (3) e do histograma, figura (3.b), concluímos que a distribuição de $\hat{\theta}_2$ é próxima da distribuição normal.

Portanto, através desse estudo de simulação, podemos verificar que o modelo (7.1) é próximo ao linear no seu comportamento de estimação, ou seja, existem evidências de que o estimador de mínimos quadrados tem um comportamento "próximo do linear".

Analisando as medidas de curvatura de Bates e Watts, temos que a medida de não-linearidade intrínseca (IN) é dada por, $IN = 0.0742$ e a medida de não-linearidade paramétrica (PE) é dada por $PE = 0.1723$.

Comparando as medidas de curvatura IN e PE com o valor $(1/2\sqrt{F}) = 0.2469$, onde $F = F_{2,10}(0.95) = 4.10$, temos que $IN < 0.2469$, o que indica que a curvatura intrínseca é adequadamente pequena e $PE < 0.2469$, indicando que a suposição de coordenadas uniformes não é rejeitada e não existe, portanto, a necessidade de uma reparametrização. Assim, podemos concluir que a curvatura é adequadamente pequena, ou seja, o grau de não-

linearidade do modelo é pequeno o suficiente para justificar o uso dos resultados assintóticos.

Através da fórmula (1.72), calculamos os vícios dos estimadores de mínimos quadrados, dados por,

$$\text{Vício } (\hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} 0.190001 \\ 0.000443 \end{pmatrix} \text{ e portanto, } \begin{cases} \% \text{Vício } (\hat{\theta}_1) = 0.0893\% \\ \% \text{Vício } (\hat{\theta}_2) = 0.6903\% \end{cases}$$

Concluimos que os vícios dos estimadores não são significativos, pois são menores que 1%. Assim, não temos evidências que os parâmetros do modelo (7.1) têm um comportamento não-linear alto, considerando o conjunto de dados da tabela (1).

As medidas de não-linearidade de Bates e Watts (1980) indicam que o grau de não-linearidade no problema de estimação não-linear, para o modelo (7.1) e conjunto de dados da tabela (1), é pequeno o suficiente para justificar a utilização dos resultados usuais da teoria assintótica. Através da fórmula de Box (1.72) para o cálculo dos vícios dos estimadores e através dos resultados obtidos da simulação, foi possível estudar o comportamento dos estimadores de mínimos quadrados e verificar que o modelo (7.1) tem comportamento próximo do linear. Finalmente, concluimos que os resultados clássicos da teoria assintótica são adequados para o problema. Assim, podemos calcular testes de hipóteses, intervalos e contornos de confiança para os parâmetros do modelo ou funções dos mesmos, baseados na teoria assintótica apresentada no Capítulo 2.

Um intervalo de confiança 95% aproximado para θ_1 usando (2.4) é dado por (197.206; 228.162), e analogamente, calculamos um intervalo de confiança 95% aproximado para θ_2 , (0.0457; 0.0825).

O gráfico do contorno de confiança 95% para a soma de quadrados residual, utilizando (1.55) é apresentado na figura (4). O contorno apresenta-se numa boa forma elíptica com uma correlação positiva, indicando que a aproximação normal para os estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ é apropriada.

Figura 4. Contorno de $S(\theta_1, \theta_2)$ para θ_1 e θ_2 .

Considerando a previsão y^* em $x^* = 1.0$, $y^* = 199.868$, um intervalo de confiança 95% aproximado para y^* é dado (de 2.8) por: (172.667; 227.069).

Como na prática, também temos interesse em tomar decisões sobre os parâmetros baseados nas informações amostrais, torna-se necessário a formulação de hipóteses estatísticas sobre os parâmetros de interesse. Desta forma, construímos testes de hipóteses utilizando a seção (2.6) do Capítulo 2.

Considerando um nível de significância $\alpha = 0.05$, vamos testar, $H_{01}:\theta_1 = 200$ versus $H_{11}:\theta_1 \neq 200$. Utilizando a estatística F_1 (ver 2.20), temos, $F_1 = 3.315$ e como $F_1 < F_{1,10}(0.05) = 4.96$ não rejeitamos a hipótese nula ao nível $\alpha = 0.05$ de significância.

Considerando as seguintes hipóteses, $H_{02}:\theta_2 = 0$ versus $H_{12}:\theta_2 \neq 0$ e utilizando a estatística F_1 (ver 2.20), temos, $F_1 = 59.968$ e ao nível $\alpha = 0.05$ de significância rejeitamos a hipótese nula $H_{02}:\theta_2 = 0$.

Análise Bayesiana

Como visto no Capítulo 4, em inferência Bayesiana, as inferências sobre os parâmetros de interesse são baseadas nas suas densidades a posteriori marginais.

Para inferir sobre os parâmetros θ_1 e θ_2 , considerando o conjunto de dados da tabela (1), construímos o gráfico da densidade a posteriori marginal para θ_1 (4.27), obtida pelo método de aproximação de Laplace, e o gráfico da densidade a posteriori marginal para θ_2 (4.32), também obtida pelo método de Laplace, ver figuras (5(a)) e (5(b)), respectivamente.

Figura 5. Densidades a Posteriori Marginais para θ_1 e θ_2 , respectivamente, obtidas pelo método de Laplace.

As modas dessas densidades a posteriori marginais são dadas por, $\tilde{\theta}_1 = 212.600$ e $\tilde{\theta}_2 = 0.063$.

No método de aproximação de Laplace, ver Apêndice D, visando melhores aproximações e simplificação dos cálculos, podemos ter várias escolhas para $f(\theta)$ em (D.2). Para o cálculo das densidades a posteriori marginais para θ_1 e θ_2 , respectivamente, escolhemos $f(\theta) = 1$. Também poderíamos considerar, para o cálculo da densidade a posteriori marginal para θ_1 (ver 4.29), $f(\theta) = \sqrt{g(\theta_2)}$. A moda a posteriori para θ_1 , nesse caso, é dada por $\tilde{\theta}_1 = 212.600$.

Figura 6. Densidade a Posteriori Marginal para θ_1 , considerando $f(\theta) = \sqrt{g(\theta_2)}$ no método de Laplace.

Das figuras (5) e (6), verificamos formas similares para as duas aproximações de Laplace, considerando o conjunto de dados da tabela (1).

Considerando a densidade a posteriori conjunta (4.24) para θ_1 e θ_2 e utilizando o método de aproximação de Laplace, calculamos, considerando a função de perda quadrática, os estimadores de Bayes para os parâmetros de interesse θ_1 e θ_2 , $E(\theta_1 / \text{dados}) = 211.384$ e $E(\theta_2 / \text{dados}) = 0.0648$, respectivamente.

Intervalos de credibilidade 95% Bayesianos para θ_1 e θ_2 , obtidos pela regra de Simpson, são dados por: $[204.1667; 220.8015]$ e $[0.0529; 0.0771]$, respectivamente.

Na figura (7), temos o gráfico da densidade preditiva (4.39) para uma observação futura em $x^* = 1$, obtida pelo método de aproximação de Laplace. A moda desta densidade preditiva é dada por $\tilde{y} = 199.900$ e um intervalo de credibilidade 95% Bayesiano para a previsão em $x^* = 1$, obtido pela regra de Simpson, é dado por, $(177.5080 \leq y \leq 224.8504)$.

Figura 7. Densidade Preditiva para uma obs. futura em $x^* = 1$.

Para verificar se o método de Laplace para aproximação de integrais tem boa precisão para essa modelagem não-linear, vamos fazer uma comparação usando o método numérico de Simpson.

Figura 8. Densidade a Posteriori Marginal para θ_1 e θ_2 , respectivamente.

As modas das densidades a posteriori marginais para θ_1 e θ_2 , obtidas pelo método numérico de Simpson, são dadas por $\tilde{\theta}_1 = 212.582$ e $\tilde{\theta}_2 = 0.0631$, respectivamente.

Observamos, através da figura (8), uma excelente concordância entre o método numérico de Simpson e o método de aproximação de Laplace, indicando que o método de Laplace tem boa precisão, apesar do tamanho amostral ser muito pequeno. É importante salientar, que a escolha de $f(\theta)$ na aproximação de Laplace (ver apêndice D) pode simplificar as expressões e fornecer resultados precisos (ver Tierney e Kadane, 1986).

Figura 9. Densidades a Posteriori Marginais para θ_1 .

Assim, através da figura (9), onde consideramos a densidade a posteriori marginal para θ_1 , observamos que existe uma grande concordância entre as densidades, considerando a aproximação dada com $f(\theta) = \sqrt{g(\theta_2)}$ e a aproximação dada com $f(\theta) = 1$.

Observamos que as densidades a posteriori marginais obtidas pelo método de aproximação de Laplace e pelo método numérico de Simpson são bem comportadas, indicando uma boa normalidade das densidades a posteriori e conseqüentemente inferências precisas.

Parâmetro	$E(\theta/dados)$ (Laplace)	$E(\theta/dados)$ (Simpson)
θ_1	211.3840	212.4347
θ_2	0.0648	0.0642

Tabela 4. Esperanças a posteriori para os parâmetros θ_1 e θ_2 obtidas das densidades a posteriori marginais através dos métodos de Laplace ((4.27) e (4.32), respectivamente) e Simpson.

Através da tabela (4), observamos que as estimativas são razoavelmente próximas, indicando que o método de aproximação de Laplace tem boa precisão.

Comparação entre as Estimativas Pontuais

Na tabela (5) temos os estimadores de máxima verossimilhança, os estimadores de mínimos quadrados e as modas a posteriori para θ_1 e θ_2 .

Parâmetro	E.M.V.	E.M.Q.	Moda Laplace	Moda Simpson
θ_1	212.684	212.684	212.600	212.582
θ_2	0.0641	0.0641	0.0630	0.0631

Tabela 5. Estimadores de máxima verossimilhança, mínimos quadrados e modas a posteriori para os parâmetros θ_1 e θ_2 .

Observamos que as estimativas pontuais para os parâmetros de interesse θ_1 e θ_2 são próximas, considerando o método clássico e o método Bayesiano com uma densidade a priori não-informativa.

Comparação entre as Estimativas por Intervalos

Para facilitar a comparação, construímos a tabela (6), onde temos inferências por intervalos para os parâmetros θ_1 e θ_2 , obtidas através das análises clássica e Bayesiana considerando uma densidade a priori não-informativa.

Normalidade assintótica de $\hat{\theta}_1$ e $\hat{\theta}_2$	I.C. cláss. obtidos através de (2.4) (ver Cap.2 - seção 2.1)	Intervalo de Credibilidade Bayesiano (Simpson)
(200.254; 225.114)	(197.206; 228.162)	(204.167; 220.802)
(0.0445; 0.0837)	(0.0457; 0.0825)	(0.0529; 0.0771)

Tabela 6. Intervalos de Confiança 95% e Intervalos de Credibilidade Bayesianos 95% para θ_1 e θ_2 .

Os intervalos para θ_1 e θ_2 são razoavelmente próximos e observamos uma amplitude menor dos intervalos para o caso Bayesiano, indicando uma precisão melhor, considerando uma priori não-informativa. Através do gráfico do contorno para θ_1 e θ_2 (ver figura (4)), observamos a existência de uma boa forma elíptica, o que sugere boa normalidade para os e.m.v. dos parâmetros θ_1 e θ_2 e sob o enfoque Bayesiano, as densidades a posteriori têm formas bem comportadas, validando assim, todos os resultados obtidos, o que pode ser confirmado através do estudo de simulação e das medidas de curvatura de Bates e Watts, que os resultados assintóticos são válidos e portanto, as inferências obtidas são precisas.

É importante salientar que poderíamos obter inferências mais precisas considerando densidades a priori informativas para os parâmetros.

O Efeito de Reparametrização nas Análises Clássica e Bayesiana

Usualmente, as análises estatísticas para os modelos não-lineares são baseadas em resultados assintóticos que dependam de uma boa parametrização. Em alguns casos, a precisão pode ser melhorada considerando uma reparametrização apropriada, o que não é necessário com os dados da tabela (1), conforme vimos no início deste capítulo. Apesar disso, vamos considerar uma reparametrização para o modelo de Michaelis-Menten (7.1) sugerida por Bates e Watts (1988),

$$\beta_1 = \frac{\theta_2}{\theta_1} \text{ e } \beta_2 = \frac{1}{\theta_1}, \quad (7.3)$$

assim, o modelo na forma reparametrizada é dado por,

$$y_i = \frac{x_i}{\beta_1 + \beta_2 x_i} + \varepsilon_i, \quad i = 1, 2, \dots, 12. \quad (7.4)$$

Os estimadores de mínimos quadrados para β_1 e β_2 , obtidos através do método numérico de Gauss-Newton, após 6 iterações (observar um número menor de iterações do que na parametrização original), são dados por,

$$\begin{aligned} \hat{\beta}_1 &= 0.0003015, & E.P.(\hat{\beta}_1) &= 0.000032 \\ \hat{\beta}_2 &= 0.004702, & E.P.(\hat{\beta}_2) &= 0.00015 \end{aligned}$$

e a soma de quadrados residual é dada por: $SQR = 1195.45$.

Um intervalo de confiança 95% aproximado usando (2.4), para os parâmetros $\hat{\beta}_1$ e $\hat{\beta}_2$, é dado por $[0.0002301; 0.0003728]$ e $[0.0043596; 0.0050440]$, respectivamente.

Assim, o modelo ajustado na forma reparametrizada é dado por,

$$\begin{aligned} \hat{y}_i &= \frac{x_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} \\ &= \frac{x_i}{0.0003015 + 0.004702 x_i} \end{aligned} \quad (7.5)$$

É importante fazer uma análise de resíduos para o modelo acima para verificar se as suposições básicas $[\varepsilon_i: i.i.d. N(0, \sigma^2)]$ estão satisfeitas.

Através da figura (10.(a)), observamos o gráfico dos resíduos contra os valores ajustados e notamos uma boa aleatorização dos pontos, indicando que a variância dos resíduos é constante. Podemos também observar que existe uma observação um pouco discrepante, mas como o número de observações ($n=12$) é pequeno, podemos considerar variância constante para os resíduos.

Figura 10. Gráfico Resíduos vs Valores Ajustados para o modelo (7.5) e gráfico Normal Probabilístico.

Através do gráfico normal probabilístico (ver figura 10(b)) observamos a existência de uma relação linear aproximada entre os "escores normais" e os resíduos; indicando uma forte tendência à normalidade dos resíduos e a existência de uma observação um pouco discrepante, que não é considerada problemática. Mas como o número de observações ($n=12$) é pequeno, podemos considerar a suposição de normalidade satisfeita para os resíduos.

O coeficiente de explicação, $R^2 = 0.961$, indica que o modelo (7.5) explica 96.1% da variabilidade dos dados. Assim, podemos concluir que o modelo (7.5) é apropriado para os dados.

Para verificar se os resultados assintóticos da teoria clássica são válidos, considerando a reparametrização (7.3), temos que analisar as medidas de curvatura de Bates e Watts, o vício de Box e realizar um estudo de simulação (ver Ratkowsky, 1983).

Considerando o modelo (7.5), as medidas de curvatura de Bates e Watts são dadas por,

- curvatura intrínseca: $IN = 0.0742$;
- curvatura paramétrica: $PE = 0.1256$.

O valor de comparação $(1/2\sqrt{F}) = 0.2469$, é maior do que as duas medidas de curvatura, indicando que a curvatura é adequadamente pequena, assim, o grau de não-linearidade do modelo (7.5) é pequeno o suficiente para garantir a validade dos resultados assintóticos.

A medida de curvatura intrínseca é a mesma que a anterior (do modelo (7.4) na parametrização original), pois essa medida não é sensível à reparametrização.

O vício de Box, calculado através de (1.72), é dado por,

$$\text{Vício } (\hat{\beta}_1, \hat{\beta}_2) = \begin{pmatrix} 0.00000116 \\ 0.00000082 \end{pmatrix}, \text{ portanto, } \begin{cases} \% \text{ Vício } (\hat{\beta}_1) = 0.3849\% \\ \% \text{ Vício } (\hat{\beta}_2) = 0.0174\% \end{cases}$$

Os vícios não são significativos, indicando que nenhum dos parâmetros possui uma alta não-linearidade.

Tendo um conjunto de resultados de simulação, é muito simples examinar o efeito de reparametrizações, pois o estimador de mínimos quadrados de uma função do parâmetro é simplesmente aquela função do estimador de mínimos quadrados do parâmetro.

Considerando a reparametrização (7.3), para avaliar o seu desempenho, devemos simplesmente transformar os 1000 estimadores de $\hat{\theta}$ em 1000 estimadores de $\hat{\beta}$ através das relações,

$$\hat{\beta}_1 = \frac{\hat{\theta}_2}{\hat{\theta}_1} \quad \text{e} \quad \hat{\beta}_2 = \frac{1}{\hat{\theta}_1}.$$

Assim, os momentos amostrais para $\hat{\beta}_1$ e $\hat{\beta}_2$ são dados pelas tabelas (7) e (8), respectivamente.

<i>Momentos Amostrais para $\hat{\beta}_1$</i>			
<i>Momentos Amostrais</i>		<i>Quantidades Derivadas</i>	
Média	0.00030378	%Vício	0.763%
Variância	0.000000001	%exc.variância	0.0000001%
3.º Momento	6.181031e-15	Skewness	0.19546
4.º Momento	3.373478e-18	exc. Curtose	0.20213

Tabela 7. Momentos Amostrais para $\hat{\beta}_1$.

<i>Momentos Amostrais para $\hat{\beta}_2$</i>			
<i>Momentos Amostrais</i>		<i>Quantidades Derivadas</i>	
Média	0.0047005	%Vício	-0.0279%
Variância	0.000000022	%exc.variância	0.0000022%
3.º Momento	5.85518e-13	Skewness	0.174597
4.º Momento	1.634682e-15	exc. Curtose	0.256540

Tabela 8. Momentos Amostrais para $\hat{\beta}_2$.

Observando o histograma, ver figura (11.a) para $\hat{\beta}_1$ e a tabela (7), concluímos que a distribuição de $\hat{\beta}_1$ é próxima da distribuição normal. Analogamente, através do histograma (figura (11b)) e dos momentos amostrais para $\hat{\beta}_2$, dados na tabela (8), concluímos que a distribuição de $\hat{\beta}_2$ também é próxima da distribuição normal.

(a)

(b)

Figura 11. Histograma para $\hat{\beta}_1$ e $\hat{\beta}_2$, respectivamente.

Assim, através das análises das medidas de curvatura, do vício de Box e do estudo de simulação, proposto por Ratkowsky (1983), podemos concluir que o modelo na forma reparametrizada (7.4) tem um grau de não-linearidade pequeno o suficiente para a utilização dos resultados assintóticos. Devemos ressaltar, que o modelo na forma reparametrizada possui um grau de não-linearidade menor que o modelo na parametrização original (7.1), pois os resultados das análises de Bates e Watts (1980), Box (1971) e Ratkowsky (1983) foram menores. Apesar disso, o modelo (7.1) pode ser utilizado na parametrização original,

pois apesar de possuir um grau de não-linearidade maior, os resultados assintóticos ainda são válidos.

Análise Bayesiana

Considerando a densidade a posteriori conjunta para θ_1 e θ_2 (ver 4.24), calculamos as densidades a posteriori marginais para β_1 (4.53) e β_2 (4.54) através do método de aproximação de integrais de Laplace (ver Apêndice D), (ver figuras (12.a) e (12.b), respectivamente).

Figura 12. Densidades a Posteriori Marginais para β_1 e β_2 , respectivamente.

A moda da densidade a posteriori marginal para β_1 (de 4.53) é dada por $\tilde{\beta}_1 = 0.0003$ e a moda a posteriori para β_2 (de 4.54) é dada por $\tilde{\beta}_2 = 0.00047$.

Através das figuras (12(a) e 12(b)) observamos claramente uma forma simétrica, bem comportada, para as densidades a posteriori marginais para β_1 e β_2 , indicando que a utilização de métodos de aproximação Bayesianos têm uma boa precisão, especialmente para aproximar momentos a posteriori.

Observando o contorno da função de verossimilhança conjunta para β_1 e β_2 (ver figura (13)), verificamos boa normalidade para a função de verossimilhança (3.8), considerando σ^2 conhecido, indicando melhoria na normalidade assintótica dos estimadores de máxima verossimilhança. Vemos ainda que, este contorno acrescenta pouca contribuição à normalidade, como podemos observar pelas figuras (4) e (13).

Figura 13. Contorno de verossimilhança para β_1 e β_2 .

Verificamos uma concordância razoável entre as inferências obtidas por intervalos e as inferências pontuais, utilizando a metodologia clássica e a Bayesiana considerando uma densidade a priori não-informativa.

As fórmulas aproximadas, usando o método de Laplace para aproximação de integrais são muito simples e não exigem métodos computacionais sofisticados. Comparando os resultados obtidos pelo método de Laplace com os obtidos por Simpson, observamos uma boa concordância entre os métodos, indicando uma boa precisão para o método de Laplace.

Considerando a parametrização original (θ_1, θ_2) , observamos que para o modelo (7.1) e o conjunto de dados da tabela (1), não existe a necessidade de reparametrizações, pois os resultados assintóticos da teoria clássica podem ser utilizados devido a baixa não-linearidade do modelo e os resultados Bayesianos aproximados também têm boa precisão.

A reparametrização (7.3), praticamente não mudou as inferências obtidas para o modelo (7.1) com os dados da tabela (1), mas dependendo do conjunto de dados considerado, as reparametrizações tornam-se necessárias.

Quando não podemos utilizar os resultados assintóticos da teoria clássica devido a alta não-linearidade do modelo e dependendo do conjunto de dados, o uso de métodos Bayesianos aproximados é uma alternativa apropriada.

No exemplo seguinte, vamos aplicar a análise Bayesiana para o modelo de Michaelis-Menten, considerando um conjunto de dados que produza uma alta curvatura paramétrica.

7.2. Exemplo 2

Neste exemplo, vamos considerar o conjunto de dados introduzido por Bates e Watts (1980), consistindo de $n = 12$ observações, dado na tabela (9). Esses dados foram simulados assumindo o modelo de Michaelis-Menten,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, i = 1, 2, \dots, 12 \quad (7.6)$$

onde x representa a variável independente referente a concentração de substrato, y a variável dependente referente a velocidade da reação, (θ_1, θ_2) são os parâmetros do modelo e vamos assumir que os erros ε_i são variáveis aleatórias $i.i.d. N(0, \sigma^2)$.

<i>Concentração de Substrato</i> (ppm) x	<i>Velocidade da reação</i> (counts/min ²) y
2.000	0.0615
2.000	0.0527
0.667	0.0334
0.667	0.0258
0.400	0.0138
0.400	0.0258
0.286	0.0129
0.286	0.0183
0.222	0.0083
0.222	0.0169
0.200	0.0129
0.200	0.0087

Tabela 9. Dados simulados do modelo de Michaelis-Menten.

Na figura (14) abaixo, temos o gráfico da velocidade da reação versus a concentração de substrato para os dados da tabela (9).

Figura 14. Gráfico da velocidade da reação versus a concentração de substrato.

Análise Clássica

Nesta seção, aplicamos aos dados da tabela (9) a metodologia de análise clássica apresentada em capítulos anteriores e estudamos o efeito de algumas reparametrizações na normalidade da função de verossimilhança, explorando o uso das medidas de Kass e Slate (1992) e de Hills e Smith (1993) introduzidas no capítulo 3.

Os estimadores de mínimos quadrados para os parâmetros θ_1, θ_2 e σ^2 do modelo (7.6), obtidos através do método de Gauss-Newton, ver seção (1.4.3), com 5 iterações são dados por,

$$\begin{aligned}\hat{\theta}_1 &= 0.1056 & E.P.(\hat{\theta}_1) &= 0.0176 \\ \hat{\theta}_2 &= 1.7027 & E.P.(\hat{\theta}_2) &= 0.4758\end{aligned}$$

$$\text{e } \hat{\sigma}^2 = 0.0000201.$$

Adequabilidade do modelo

Considerando o modelo de Michaelis-Menten (7.6) e os dados da tabela (9), o modelo ajustado por mínimos quadrados pelo método de Gauss-Newton é dado por,

$$\hat{y}_i = \frac{0.1056x_i}{1.7027 + x_i}, i = 1, 2, \dots, 12 \quad (7.7)$$

onde as suposições básicas são: $\varepsilon_i: i.i.d. N(0, \sigma^2)$.

Através da figura (15.a), observamos que os pontos estão aleatoriamente distribuídos em torno de zero, indicando que a variância é constante e a média próxima de zero. O gráfico normal probabilístico, informa-nos que existe uma normalidade aproximada para os resíduos. Ressaltando que o número de observações é pequeno, podemos aceitar a normalidade dos resíduos. O modelo explica 93.8% da variabilidade dos dados, o que é muito satisfatório e assim, concluímos que o modelo (7.7) é satisfatório.

Figura 15. Gráfico dos resíduos do modelo (7.7) versus os valores ajustados e Gráfico normal probabilístico, respectivamente.

Os estimadores de máxima verossimilhança para os parâmetros θ_1, θ_2 e σ^2 , obtidos das equações (2.30) são dados por, $\hat{\theta}_1 = 0.1056$ e $\hat{\theta}_2 = 1.7027$ e $\hat{\sigma}^2 = 0.00001675$. A matriz de informação de Fisher (2.33) para θ_1, θ_2 e σ^2 calculada em $(\hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2)$ é dada por,

$$I(\theta_1, \theta_2, \sigma^2) = \begin{pmatrix} 53980 & -1924 & 0 \\ -1924 & 74 & 0 \\ 0 & 0 & 21373546496 \end{pmatrix}$$

A medida de não-linearidade intrínseca é dada por $IN = 0.0650$ e a não-linearidade paramétrica por $PE = 0.5976$. Comparando estes resultados com $\frac{1}{2\sqrt{F}} = 0.2469$, onde $F = F_{2,10}(0.95) = 4.10$, observamos que $IN < 0.2469$, isto é, a não-linearidade intrínseca do locus da solução é adequadamente pequena, mas a curvatura paramétrica, $PE > 0.2469$, não é adequadamente pequena (ver por exemplo, Ratkowsky, 1983), sendo necessário a utilização de uma boa reparametrização.

Os vícios dos estimadores são dados por,

$$\text{Vício}(\hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} 0.3125\text{E} - 02 \\ 0.9358\text{E} - 01 \end{pmatrix} \text{ e portanto, } \begin{cases} \% \text{vício}(\hat{\theta}_1) = 2.9577\% \\ \% \text{vício}(\hat{\theta}_2) = 5.4958\% \end{cases}$$

Como essas porcentagens são maiores que 1%, temos um indicativo de comportamento não-linear (ver Ratkowsky, 1983).

Na figura (17), temos o contorno da função de verossimilhança conjunta para (θ_1, θ_2) , assumindo σ^2 conhecido. Observamos claramente, que este contorno não tem boa forma elíptica indicando a necessidade de uma reparametrização para melhorar a normalidade da função de verossimilhança, e assim, melhorar a precisão dos resultados assintóticos. O que pode ser confirmado pelas medidas de curvatura e pelas medidas de vício dadas.

Assim, como na parametrização (θ_1, θ_2) temos valores grandes para a medida de curvatura paramétrica e grandes porcentagens para os vícios dos estimadores, devemos considerar uma reparametrização para obter melhores inferências aproximadas para os parâmetros do modelo (7.6).

Figura 16. Função de Verossimilhança conjunta para (θ_1, θ_2) .

(a) Aproximado

(b) Exato (Maple)

Figura 17. Contorno da função de verossimilhança de θ_1 e θ_2 (σ^2 conhecido).

Considerando a reparametrização proposta por Bates e Watts (1988), $\beta_1 = \theta_2 / \theta_1$; $\beta_2 = 1 / \theta_1$, encontramos os estimadores de máxima verossimilhança dados por $\hat{\beta}_1 = 16.1174$, $\hat{\beta}_2 = 9.4659$ e $\hat{\sigma}^2 = 0.0000167$. Os estimadores de mínimos quadrados para os parâmetros β_1 , β_2 e σ^2 , obtidos pelo método de Gauss-Newton, com 4 iterações, são dados por, $\hat{\beta}_1 = 16.1174$, $\hat{\beta}_2 = 9.4659$ e $\hat{\sigma}^2 = 0.0000201$.

Na figura (19) temos o contorno da função de verossimilhança conjunta para (β_1, β_2) , assumindo σ^2 conhecido, apresentando-se de forma elíptica, indicando um comportamento normal da função de verossimilhança, ou seja, temos uma melhor normalidade para a função de verossimilhança conjunta (figura 18).

Figura 18. Função de Verossimilhança conjunta para (β_1, β_2) .

As medidas de curvatura de Bates e Watts (1980) para o modelo de Michaelis-Menten (7.6) na parametrização (β_1, β_2) são dadas por: $IN = 0.0650$ e $PE = 0.2259$, indicando uma redução significativa na curvatura paramétrica ($PE < 0.2469$). A medida do vício de Box (1971) é dada por,

$$\text{Vício}(\hat{\beta}_1, \hat{\beta}_2) = \begin{pmatrix} 0.1335 \\ -0.1725 \end{pmatrix} \text{ e portanto, } \begin{cases} \% \text{Vício}(\hat{\beta}_1) = 0.8285\% \\ \% \text{Vício}(\hat{\beta}_2) = -0.1822\% \end{cases}$$

e são adequadamente pequenos ($< 1\%$).

Figura 19. Contorno da função de verossimilhança de β_1 e β_2 (σ^2 conhecido).

Considerando a parametrização $\delta_1 = \theta_1 u / (\theta_2 + u)$ e $\delta_2 = \theta_1 v / (\theta_2 + v)$ com $u = 0.396$ e $v = 1.945$ (ver capítulo 3), os estimadores de máxima verossimilhança para os parâmetros δ_1 , δ_2 e σ^2 são dados por: $\hat{\delta}_1 = 0.0199$, $\hat{\delta}_2 = 0.0563$ e $\hat{\sigma}^2 = 0.0000167$. Os estimadores de mínimos quadrados, obtidos pelo método de Gauss-Newton com 4 iterações, são dados por: $\hat{\delta}_1 = 0.0199$, $\hat{\delta}_2 = 0.0563$ e $\hat{\sigma}^2 = 0.0000201$.

Na figura (21) temos o gráfico do contorno da função de verossimilhança conjunta de (δ_1, δ_2) assumindo σ^2 conhecido, a qual é mostrada na figura (20). Observamos claramente a forma elíptica com a qual se apresenta o contorno, indicando uma melhor normalidade da função de verossimilhança.

As medidas de curvatura de Bates e Watts (1980) para o modelo de Michaelis-Menten nesta parametrização são dadas por: $IN = 0.0650$ e $PE = 0.0002$ e como essas medidas são menores que o valor de comparação ($IN, PE < 0.2469$), a curvatura IN é adequadamente pequena e a suposição de coordenadas uniformes não é rejeitada, indicando que o grau de

não-linearidade do modelo é pequeno o suficiente para justificar os resultados assintóticos da teoria clássica.

Figura 20. Função de Verossimilhança conjunta para (δ_1, δ_2) .

Os vícios dos estimadores são dados por,

$$\text{Vício}(\hat{\delta}_1, \hat{\delta}_2) = \begin{pmatrix} -0.9617\text{E} - 07 \\ 0.4124\text{E} - 07 \end{pmatrix} \text{ e portanto, } \begin{cases} \% \text{Vício}(\hat{\delta}_1) = -0.0005\% \\ \% \text{Vício}(\hat{\delta}_2) = 0.0001\% \end{cases}$$

Assim, através dessas medidas observamos uma melhora significativa da função de verossimilhança e do comportamento não-linear do modelo, especialmente, na parametrização (δ_1, δ_2) .

Figura 21. Contorno da função de verossimilhança de δ_1 e δ_2 (σ^2 conhecido).

Uma outra forma de se avaliar a normalidade da função de verossimilhança é através da função de verossimilhança *profile* e do *T-Plot* de Hills e Smith (1993), para θ_1 e θ_2 apresentados nas figuras (23) e (22), respectivamente. Observamos claramente formas curvilíneas para os *T-Plots* na parametrização (θ_1, θ_2) , indicando a não-normalidade das funções de verossimilhança *profile*.

Figura 22. *T-Plots das Verossimilhanças Profile de θ_1 e θ_2 , respectivamente.*

Figura 23. *Função de Verossimilhança Profile de θ_1 e θ_2 , respectivamente.*

Na figura (24), temos os *T-Plots* para as funções de verossimilhança *profile* (figura(25)) de β_1 e β_2 . Nesta parametrização houve uma melhoria da linearidade dos *T-Plots*, mas ainda observamos gráficos curvos.

Figura 24. T-Plots das Verossimilhanças Profile de β_1 e β_2 , respectivamente.

Figura 25. Função de Verossimilhança Profile de β_1 e β_2 , respectivamente.

Na figura (26) temos os T-Plots para as funções de verossimilhança *profile* para δ_1 e δ_2 (figura (27)). Nesta parametrização, os gráficos apresentam boa linearidade, indicando a normalidade das funções de verossimilhança *profile* apresentadas na figura (27).

Figura 26. T-Plots das Verossimilhanças Profile de δ_1 e δ_2 , respectivamente.

Figura 27. Função de Verossimilhança Profile de δ_1 e δ_2 , respectivamente.

Para verificar a normalidade conjunta da função de verossimilhança de (θ_1, θ_2) , podemos usar a terceira derivada sumária para diagnosticar a contribuição global de normalidade (Kass e Slate, 1992), dada de (3.7) por $m^2 \bar{B}^2 = 1.8924$. Na parametrização (β_1, β_2) obtemos $m^2 \bar{B}^2 = 0.2788$ e na parametrização (δ_1, δ_2) encontramos $m^2 \bar{B}^2 = 0.0000$, indicando uma grande melhoria da normalidade conjunta da função de verossimilhança. Esses resultados obtidos da derivada terceira sumária confirmam os obtidos através dos gráficos de contorno.

Assim, através das medidas de não-normalidade de Kass e Slate (1992), dos gráficos de contorno e dos *T-Plots*, observamos a eficiência das reparametrizações (3.1) e (3.2) na melhoria da normalidade da função de verossimilhança, com especial destaque para a parametrização (δ_1, δ_2) que fornece melhor precisão nos resultados assintóticos.

Sob este ponto de vista, observamos que é possível obter inferências clássicas precisas considerando a parametrização (δ_1, δ_2) , pois o objetivo maior é a previsão. Mas, em trabalhos práticos, usualmente, os pesquisadores gostariam de obter inferências precisas para os parâmetros originais θ_1 e θ_2 , que têm interpretações físicas especiais, ver seção (7.1.1) e nestas situações, uma análise Bayesiana pode ser uma alternativa apropriada para a obtenção de inferências precisas para os parâmetros originais θ_1 e θ_2 .

Análise Bayesiana

Como vimos anteriormente, na metodologia Bayesiana, as inferências sobre os parâmetros de interesse são baseadas nas suas densidades a posteriori marginais, resultando em um procedimento que pode simplificar as interpretações e conclusões para vários problemas práticos.

Considerando a densidade a priori não-informativa (4.21) e os dados da tabela (9), temos, na figura (28), os gráficos das densidades a posteriori marginais para θ_1 e θ_2 . As modas são dadas por: $\tilde{\theta}_1 = 0.103$ e $\tilde{\theta}_2 = 1.609$, usando o método de Laplace. Usando a regra de Simpson, temos, $\tilde{\theta}_1 = 0.101$ e $\tilde{\theta}_2 = 1.560$.

Figura 28. Gráfico das densidades a posteriori marginais para θ_1 e θ_2 , respectivamente.

Considerando a função de perda quadrática dada em (4.14), uma outra estimativa Bayesiana para θ_1 e θ_2 obtida de (4.35) e (4.36) é dada por: $E(\theta_1 / \text{dados}) = 0.106$ e $E(\theta_2 / \text{dados}) = 1.712$. Usando a regra de Simpson, obtemos $E(\theta_1 / \text{dados}) = 0.106$ e $E(\theta_2 / \text{dados}) = 1.724$. Assim, observamos que os resultados Bayesianos aproximados por Laplace têm boa precisão.

Usando a regra de Simpson, obtemos de (4.27) e (4.32), intervalos 95% de credibilidade Bayesianos para θ_1 e θ_2 dados por: (0.09712; 0.11408) e (1.44372; 2.01267), respectivamente.

Na figura (29), temos o gráfico da densidade preditiva para uma observação futura y^* (4.39) em um dado valor fixo $x^* = 1$ da variável independente no modelo (7.6). A moda dessa densidade é dada por $\tilde{y} = 0.0390$ e um intervalo 95% de credibilidade Bayesiano, obtido através do método numérico de Simpson, é dado por (0.02910; 0.04952).

Um outro estimador pontual Bayesiano, considerando a função de perda quadrática, é dado por $E_{\theta_1, \theta_2}(y^* / \text{dados}) = 0.0389$.

Figura 29. Densidade Preditiva p/ uma observação futura em $x^* = 1$.

Considerando a reparametrização proposta por Bates e Watts (3.1), temos, na figura (30) as densidades a posteriori marginais para β_1 e β_2 , obtidas pelo método de Laplace. As modas destas densidades são dadas por: $\tilde{\beta}_1 = 15.843$ e $\tilde{\beta}_2 = 9.442$.

Figura 30. Gráficos das densidades a Posteriori marginais para β_1 e β_2 , respectivamente.

A densidade preditiva para uma observação futura y^* em $x^* = 1$ é apresentada na figura (31) e a moda dessa densidade é dada por $\tilde{y} = 0.0390$. Usando a regra de Simpson, uma outra estimativa Bayesiana para y^* , obtida pela esperança a posteriori é dada por $E_{\beta_1, \beta_2}(y^* / \text{dados}) = 0.0387$.

Figura 31. Densidade Preditiva p/ uma observação futura em $x^* = 1$.

Na figura (32), temos as densidades a posteriori marginais para δ_1 e δ_2 , obtidas através do método numérico de Simpson. As modas são dadas por $\tilde{\delta}_1 = 0.0199$ e $\tilde{\delta}_2 = 0.0564$.

Uma outra estimativa Bayesiana para os parâmetros δ_1 e δ_2 , usando a regra de Simpson, é dada por $E(\delta_1 / \text{dados}) = 0.0198$ e $E(\delta_2 / \text{dados}) = 0.0564$. Na figura (33), usando a regra de Simpson, observamos o gráfico da densidade preditiva para uma observação futura em $x^* = 1$; a moda dessa densidade é dada por $\tilde{y} = 0.0389$ e uma outra estimativa pontual Bayesiana é dada por $E_{\delta_1, \delta_2}(y^* / \text{dados}) = 0.0387$.

Figura 32. Gráficos das densidades a Posteriori marginais para δ_1 e δ_2 , respectivamente.

Figura 33. Densidade Preditiva p/ uma observação futura em $x^* = 1$.

Na tabela (10) apresentamos as estimativas Bayesianas para uma observação futura em $x^* = 1$, considerando o modelo de Michaelis-Menten (7.6) e as três parametrizações em estudo. Comparando os valores encontrados, podemos verificar claramente que as parametrizações acrescentam pouca contribuição em termos de precisão para a densidade preditiva.

y^*	Parametrização (θ_1, θ_2)	(β_1, β_2)	(δ_1, δ_2)
Moda a Posteriori	0.0390	0.0390	0.0389
Esperança a Posteriori	0.0390	0.0387	0.0387

Tabela 10. Estimativas pontuais para a previsão em $x^* = 1$, considerando as três parametrizações em estudo.

Na prática, os pesquisadores podem ter opiniões a priori sobre os parâmetros do modelo. Isto implica que quanto maior a precisão a priori, mais fechada será a distribuição a posteriori dos parâmetros e consequentemente inferências mais precisas podem ser obtidas através dos métodos Bayesianos.

Considerando o modelo de Michaelis-Menten e o conjunto de dados da tabela (9), apresentamos, a seguir, os gráficos das densidades a posteriori marginais para θ_1 e θ_2 , figuras (34) e (35), respectivamente, nas quais PNI representa a densidade a priori não-informativa (4.21) e P_i , $i = 1, \dots, 4$ representa a densidade a priori informativa (4.41).

Na figura (34) temos os gráficos das densidades a posteriori marginais para θ_1 . Para a densidade (4.45) consideramos diferentes valores para a_1 , b_1 , σ_1 , σ_2 e ρ . Esta densidade é representada por linhas contínuas nos quatro casos considerados e a densidade (4.27) é representada por uma linha tracejada. Analogamente, na figura (35) temos os gráficos das densidades a posteriori marginais para θ_2 , onde a densidade (4.32) é representada por uma linha tracejada e a densidade (4.47) representada por linhas contínuas, onde consideramos diferentes valores para a_1 , b_1 , σ_1 , σ_2 e ρ .

Os valores considerados para a_1 , b_1 , σ_1 , σ_2 e ρ são apresentados na tabela (11). Em ambos os casos, representamos estas densidades por P_1, P_2, P_3 e P_4 .

Densidade a Posteriori	a_1	b_1	σ_1	σ_2	ρ
P_1	0.105	1.703	0.0176	0.476	0.967
P_2	0.105	1.703	0.0176	0.476	0.000
P_3	0.100	1.600	0.0176	0.476	0.967
P_4	0.105	1.703	0.0300	0.600	0.940

Tabela 11. Valores considerados para a_1 , b_1 , σ_1 , σ_2 e ρ .

Nas figuras (34) e (35) observamos que as densidades (4.45) e (4.47) apresentam-se de forma mais fechada comparadas com a densidade a posteriori (4.27) e (4.32), implicando em melhor precisão nas inferências, quando utilizamos densidades a priori informativas.

Figura 34. Gráfico da densidade a posteriori marginal para θ_1 .

Figura 35. Gráfico da densidade a posteriori marginal para θ_2 .

Na tabela (12) apresentamos estimativas Bayesianas para os parâmetros θ_1 e θ_2 do modelo de Michaelis-Menten, considerando densidades a priori informativas e a densidade a priori não-informativa (4.21).

Densidade a Posteriori	Moda a Posteriori para θ_1	Moda a Posteriori para θ_2
P_1	0.106	1.720
P_2	0.106	1.720
P_3	0.104	1.670
P_4	0.106	1.750
PNI	0.103	1.609

Tabela 12. Estimativas Bayesianas para os parâmetros do modelo de Michaelis-Menten, considerando densidades a priori informativas e não-informativa.

Concluimos, portanto, que obtemos resultados Bayesianos mais precisos quando utilizamos densidades a priori informativas apropriadas.

7.3. Exemplo 3

Neste exemplo, geramos $n = 30$ observações y_i , a variável dependente do modelo de Michaelis-Menten (7.6), com $\theta_1 = 6$ e $\theta_2 = 0.5$, tal que os erros ε_i tenham distribuição lognormal, isto é, $\log_e \varepsilon_i \sim N(0,1)$. Os dados são apresentados na tabela (13).

Conc. de Substrato x	Veloc. da Reação y	Conc. de Substrato x	Veloc. da Reação y	Conc. de Substrato x	Veloc. da Reação y
0.5	52.91	3.0	436.56	5.5	580.61
0.5	19.59	3.0	32.76	6.0	64.59
0.5	8.18	4.0	26.18	6.0	132.03
1.0	59.31	4.0	118.73	6.0	109.66
1.0	76.76	4.0	1474.82	7.0	48.88
2.0	173.98	4.0	120.67	7.0	42.25
2.0	139.95	5.0	23.67	7.5	502.39
2.0	498.67	5.0	122.98	7.5	313.84
2.5	222.73	5.0	205.40	8.0	305.84
2.5	142.26	5.5	257.97	8.0	201.10

Tabela 13. Dados gerados do modelo de Michaelis-Menten.

Dessa forma, surge uma questão prática: é possível transformar y tal que o modelo de Michaelis-Menten, com erros normais seja adequado?

Considerando o conjunto de dados da tabela (13), sem transformação e o modelo de Michaelis-Menten, os estimadores de mínimos quadrados para os parâmetros θ_1, θ_2 e σ^2 , obtidos pelo método de Gauss-Newton (seção 1.4.3) via SAS, após 44 iterações, são dados por, $\hat{\theta}_1 = 306.7175$, $E.P.(\hat{\theta}_1) = 155.3273$; $\hat{\theta}_2 = 1.1334$, $E.P.(\hat{\theta}_2) = 2.3849$ e $\hat{\sigma}^2 = 79272.27$.

O modelo ajustado é dado por,

$$\hat{y}_i = \frac{306.7175x_i}{1.1334 + x_i}, \quad i = 1, 2, \dots, 30 \quad (7.8)$$

Observando o gráfico (b) da figura (36), verificamos que os pontos não estão distribuídos aleatoriamente, existe uma tendência sistemática da variância, indicando

heterocedasticidade da mesma. O gráfico normal probabilístico, figura (36.b), nos indica a não-normalidade dos resíduos, pois o gráfico apresenta-se de forma curva.

Figura 36. Gráfico Normal Probabilístico e Gráfico dos Resíduos versus valores ajustados, respectivamente: dados da tabela (13).

Através do coeficiente de explicação, $R^2 = 0.409$, observamos que o modelo ajustado explica 40.9% da variabilidade dos dados, o que é insatisfatório. Assim, podemos concluir, que o modelo ajustado (7.8) não é adequado para os dados.

As medidas do vício de Box (1971) para os estimadores de mínimos quadrados, são dados por,

$$\text{Vício}(\hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} 118.3297 \\ 2.3027 \end{pmatrix} \text{ e portanto } \begin{cases} \% \text{Vício}(\hat{\theta}_1) = 38.5794\% \\ \% \text{Vício}(\hat{\theta}_2) = 203.1671\% \end{cases}$$

indicando uma grande discrepância entre os estimadores dos parâmetros e seus verdadeiros valores. Observamos na figura (37) o gráfico da função de verossimilhança conjunta para (θ_1, θ_2) e através da medida de não-normalidade conjunta de Kass e Slate, $m^2 \bar{B}^2 = 109.2568$, concluímos pela não-normalidade da função de verossimilhança.

Figura 37. Gráfico da função de verossimilhança conjunta de (θ_1, θ_2) .

Dessa forma, transformamos a variável dependente y para que o modelo Michaelis-Menten com erros normais seja adequado. Consideramos a família de transformação potência (Box e Cox, 1964) para a variável dependente,

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \ln y & \text{se } \lambda = 0 \end{cases} \quad (7.9)$$

tal que $y^{(\lambda)}$ tenha distribuição normal, como apresentamos no Capítulo 6.

A densidade a posteriori marginal (6.11) para λ , na transformação de Box e Cox (7.9) obtida pela regra de Simpson é apresentada na figura (38). A moda desta densidade é dada por $\tilde{\lambda} = 0.0$.

Figura 38. Gráfico da Densidade a Posteriori Marginal para λ .

Na tabela (14) temos os dados transformados, o valor de λ na transformação é igual à moda de $p(\lambda / \text{dados})$ e assim, reestimamos os parâmetros θ_1, θ_2 e σ^2 , através do método de Gauss-Newton, com 9 iterações, dados por, $\hat{\theta}_1 = 5.3789$, $E.P.(\hat{\theta}_1) = 0.3148$; $\hat{\theta}_2 = 0.3017$, $E.P.(\hat{\theta}_2) = 0.1474$ e $\hat{\sigma}^2 = 1.0364$.

Conc. de Substrato x	Veloc. da Reação y	Conc. de Substrato x	Veloc. da Reação y	Conc. de Substrato x	Veloc. da Reação y
0.5	3.96859	3.0	6.07893	5.5	6.36408
0.5	2.97502	3.0	3.48921	6.0	4.16806
0.5	2.10169	4.0	3.26500	6.0	4.88303
1.0	4.08278	4.0	4.77685	6.0	4.69738
1.0	4.34068	4.0	7.29629	7.0	3.88937
2.0	5.15894	4.0	4.79306	7.0	3.74360
2.0	4.94129	5.0	3.16421	7.5	6.21938
2.0	6.21194	5.0	4.81202	7.5	5.74888
2.5	5.40596	5.0	5.32496	8.0	5.72306
2.5	4.95766	5.5	5.55284	8.0	5.30380

Tabela 14. Dados transformados segundo a família de transformação de potência (6.1).

Os estimadores de máxima verossimilhança obtidos pelo método de Newton-Raphson, ver equações (2.30), são dados por: $\hat{\theta}_1 = 5.3789$, $\hat{\theta}_2 = 0.3017$ e $\hat{\sigma}^2 = 0.9673$. A matriz de informação de Fisher para θ_1, θ_2 e σ^2 calculada em $(\hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2)$ é dada de (2.33) por,

$$I(\theta_1, \theta_2, \sigma^2) = \begin{pmatrix} 23.1994 & -37.2443 & 0 \\ -37.2443 & 105.800 & 0 \\ 0 & 0 & 13.9647 \end{pmatrix}$$

Considerando a normalidade assintótica dos e.m.v. construímos os intervalos de confiança 95% para os parâmetros θ_1, θ_2 e σ^2 , dados na tabela (14). Em trabalhos práticos é comum considerar a matriz de informação observada I_o em lugar da matriz de informação de Fisher.

$$I_o(\theta_1, \theta_2, \sigma^2) = \begin{pmatrix} 23.1994 & -37.2443 & -0.000003 \\ -37.2443 & 111.335 & 0.000004 \\ -0.000003 & 0.000004 & 12.1027 \end{pmatrix}$$

Parâmetros	Informação de Fisher $I(\theta_1, \theta_2, \sigma^2)$	Informação Observada $I_o(\theta_1, \theta_2, \sigma^2)$
θ_1	(4.7618; 5.9959)	(4.7808; 5.9769)
θ_2	(0.0127; 0.5907)	(0.0287; 0.5747)
σ^2	(0.5119; 1.5609)	(0.4730; 1.5998)

Tabela 15. Intervalos 95% de confiança para os parâmetros.

Adequabilidade do modelo

O modelo ajustado com os dados transformados, é dado por,

$$\hat{y}_i = \frac{5.3789x_i}{0.3017 + x_i}, \quad i = 1, 2, \dots, 30 \quad (7.10)$$

O gráfico normal probabilístico, figura (39.a), apresenta-se com uma boa linearidade, indicando normalidade para os resíduos.

Na figura (39.b), observamos uma aleatorização dos pontos em torno de zero, indicando homocedasticidade da variância dos resíduos. O modelo explica 96% da variabilidade dos dados, o que é satisfatório. Portanto, concluímos que o modelo ajustado (7.10) é adequado aos dados da tabela (14).

Figura 39. Gráfico Normal Probabilístico e gráfico dos resíduos versus valores ajustados, respectivamente: dados da tabela (14).

Os vícios dos estimadores são dados por,

$$\text{Vício}(\hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} 0.2534\text{E} - 01 \\ 0.2077\text{E} - 01 \end{pmatrix} \text{ e portanto, } \begin{cases} \% \text{Vício}(\hat{\theta}_1) = 0.4708\% \\ \% \text{Vício}(\hat{\theta}_2) = 6.8836\% \end{cases}$$

Observamos um decréscimo significativo para os vícios dos estimadores, mas o vício do estimador de θ_2 ainda é significativo ($>1\%$). As medidas de curvatura de Bates e Watts (1980) são dadas por: $IN = 0.1404$ e $PE = 0.4246$. Verificamos que a curvatura paramétrica é significativa $\left(PE > \frac{1}{2\sqrt{F}} = 0.2736, \text{ onde } F = F_{2,28}(0.95) = 3.34 \right)$. Dessa forma, devemos utilizar uma reparametrização com a finalidade de melhorar os resultados assintóticos da teoria clássica.

Na figura (40), observamos o gráfico da função de verossimilhança conjunta de (θ_1, θ_2) . A medida de não-normalidade conjunta de Kass e Slate para esta função de verossimilhança é dada por $m^2 \bar{B}^2 = 1.0596$, que apesar de ser bem menor que a anterior, não é próxima de zero, indicando a necessidade de uma reparametrização.

Portanto, após a transformação da variável dependente y_i , podemos ajustar o modelo de Michaelis-Menten com erros normais, mas, para a obtenção de resultados assintóticos mais precisos, necessitamos de uma reparametrização adequada ou como alternativa, utilizamos a metodologia Bayesiana para inferir sobre os parâmetros originais.

Figura 40. Gráfico da função de verossimilhança conjunta de (θ_1, θ_2) .

Análise Bayesiana

Na metodologia Bayesiana, as análises são baseadas nas densidades a posteriori marginais dos parâmetros de interesse. Assim, com os dados transformados (tabela (14)), construímos os gráficos das densidades a posteriori marginais para θ_1 e θ_2 , considerando a densidade a priori não-informativa (4.21) e utilizando a regra de Simpson.

Figura 41. Densidades a Posteriori Marginais p/θ_1 e θ_2 , respectivamente.

A moda das densidades a posteriori marginais (4.27) e (4.30), considerando os dados da tabela (14), são dadas por: $\tilde{\theta}_1 = 5.360$ e $\tilde{\theta}_2 = 0.267$. Uma outra estimativa Bayesiana, considerando a função de perda quadrática é dada, pela regra de Simpson, por: $E(\theta_1/dados) = 5.405$ e $E(\theta_2/dados) = 0.322$. Usando a regra de Simpson, um intervalo de

credibilidade 95% Bayesiano para os parâmetros θ_1 e θ_2 são dados por (5.1719;5.5814) e (0.1834;0.4372), respectivamente.

A densidade preditiva para uma observação futura y^* , calculada em um valor fixo da variável dependente $x^* = 1$ é apresentada na figura (42). A moda é dada por $\tilde{y} = 4.123$ e uma outra estimativa Bayesiana para os parâmetros θ_1 e θ_2 , obtida pelo método numérico de Simpson, é dada por $E(y / \text{dados}) = 4.118$.

Figura 42. Gráfico da Densidade Preditiva p/ uma observação futura em $x^* = 1$.

Na figura (43), temos os gráficos das densidades a posteriori marginais para θ_1 e θ_2 , considerando a densidade a priori informativa (4.40). Os valores considerados para $a_1, b_1, \sigma_1, \sigma_2$ e ρ são dados na tabela (16).

Densidade a Posteriori	a_1	b_1	σ_1	σ_2	ρ
P_1	5.38	0.30	0.3148	0.1474	0.000
P_2	5.38	0.30	0.3148	0.1474	0.752

Tabela 16. Valores considerados para $a_1, b_1, \sigma_1, \sigma_2$ e ρ .

Figura 43. Densidades a Posteriori Marginais para θ_1 e θ_2 , respectivamente, considerando densidades a priori informativas.

A moda a posteriori para θ_1 e θ_2 são apresentadas na tabela (17). Quando consideramos a informação a priori, as inferências são precisas; assim, o procedimento Bayesiano torna-se uma ferramenta útil e poderosa na resolução deste problema.

<i>Parâmetro</i>	<i>Moda a Posteriori</i> P_1	<i>Moda a Posteriori</i> P_2	<i>Moda a Posteriori</i> PNI
θ_1	5.392	5.385	5.360
θ_2	0.300	0.304	0.267

Tabela 17. Modas a Posteriori para os parâmetros θ_1 e θ_2 .

Concluimos que, através da análise Bayesiana, podemos inferir sobre os parâmetros originais do modelo de Michaelis-Menten, considerando densidades a priori informativas apropriadas, que forneçam resultados precisos.

7.4. Exemplo 4

Neste exemplo, consideramos dois conjuntos de dados referentes a um experimento relacionando a velocidade inicial y de uma reação química com a concentração x de substrato, com o qual o experimento foi conduzido. A dependência de y em x é representada pelo modelo não-linear de Michaelis-Menten,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, \quad (7.11)$$

onde θ_1 e θ_2 são os parâmetros a serem estimados e o modelo (7.11), neste contexto, é conhecido como equação de Michaelis-Menten.

Desejamos estudar o efeito de não-normalidade nas inferências do modelo de Michaelis-Menten (7.11), supondo que os erros ε_i são variáveis aleatórias independentes com distribuição de potência exponencial dada em (5.3), através da metodologia Bayesiana.

7.4.1. Análise Bayesiana considerando β conhecido

Inicialmente, consideramos o conjunto de dados da tabela (1) com $n=12$ observações. De (5.8) a densidade a posteriori conjunta para (θ_1, θ_2) , considerando β conhecido é dada por,

$$p(\theta_1, \theta_2 / \beta, \text{dados}) \propto [M(\theta_1, \theta_2)]^{-6(1+\beta)} \quad (7.12)$$

onde $0 < \theta_1, \theta_2 < \infty$ e $M(\theta_1, \theta_2) = \sum_{i=1}^{12} \left| y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right|^{2/(1+\beta)}$.

Na tabela (18), temos a moda da densidade (7.12), $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, para cada valor fixo de β .

β	θ_1	θ_2
-0.90	200.620	0.0478
-0.85	201.204	0.0482
-0.75	202.309	0.0493
-0.65	203.438	0.0506
-0.50	205.348	0.0531
-0.40	206.834	0.0553
-0.30	208.503	0.0578
-0.25	209.358	0.0592
-0.20	210.184	0.0604
-0.10	211.671	0.0625
0.00	212.684	0.0641
0.10	213.316	0.0653
0.20	213.716	0.0663
0.25	213.871	0.0668
0.30	214.011	0.0672
0.40	214.298	0.0682
0.50	214.675	0.0694
0.65	215.239	0.0711
0.75	215.447	0.0717
0.85	215.448	0.0717
0.90	215.448	0.0717

Tabela 18. Valores da moda de θ como função de β : usando os dados da tabela (1).

A figura (44) mostra, para cada valor de β considerado, o contorno da correspondente densidade a posteriori de (θ_1, θ_2) , (7.12), calculado na moda $(\hat{\theta}_1, \hat{\theta}_2)$, os quais foram obtidos pelo software "Maple".

(i) $\beta = -0.90$

(ii) $\beta = -0.60$

(iii) $\beta = -0.30$

(iv) $\beta = 0.00$

(v) $\beta = 0.30$

(vi) $\beta = 0.60$

(vii) $\beta = 0.90$

Figura 44.(i-vii) Contornos da densidade a posteriori conjunta de (θ_1, θ_2) para vários valores de β .

(i) $\beta = -0.90$

(ii) $\beta = 0.00$

(iii) $\beta = 0.60$

Figura 45.(i-iii) Densidades a posteriori conjunta de (θ_1, θ_2) para vários valores de β .

Os contornos, juntamente com os valores das modas a posteriori dados na tabela (18), mostram como as inferências sobre (θ_1, θ_2) são afetadas pelos valores de β considerados.

Para este conjunto de dados, tanto a forma como a posição dos contornos mudam sensivelmente quando β decresce, especialmente para valores inferiores a -0.3. Em particular, os parâmetros θ_1 e θ_2 tornam-se cada vez mais correlacionados positivamente e a dispersão não está se alterando significativamente, quando β decresce a partir de -0.3, assim, a distribuição torna-se mais e mais platicúrtica.

Com os valores das modas $(\hat{\theta}_1, \hat{\theta}_2)$ obtidos na tabela (18), temos as curvas ajustadas $\hat{y}_i = \hat{\theta}_1 x_i / (\hat{\theta}_2 + x_i)$ versus x_i na figura (46), para os valores de β considerados. Observamos, claramente, que as curvas ajustadas diferem entre si para os diferentes valores de β considerados, indicando uma grande sensibilidade das inferências considerando os diferentes valores de β .

Figura 46. Curvas ajustadas para os valores de β considerados na tabela (18): dados tabela (1).

Considerando diferentes valores para β , na figura (47), temos os gráficos normais probabilísticos. Observamos uma boa linearidade no gráfico (v) para $\beta = 0$. Para os valores de $\beta < 0$, observamos, claramente, formas curvilíneas para os gráficos, especialmente para o gráfico (i) da figura (47). Para os valores $\beta > 0$, os gráficos apresentam-se com formas marcadamente curvilíneas, mas com um agravante: mostra, claramente, a existência de uma observação um pouco discrepante.

Figura 47.(i-ix). Gráficos Normais Probabilísticos.

Na figura (48), temos os gráficos dos resíduos *versus* os valores ajustados, para os diferentes valores de β considerados. Embora o número de observações seja pequeno, não observamos a aleatorização dos pontos, exceto para os valores de β compreendidos entre -0.20 a 0.20 , onde temos a aleatorização dos pontos, sendo mais evidente para $\beta = 0$. Existem, para os demais gráficos certas tendências, indicando a heterocedasticidade da variância.

Figura 48.(i-ix). Gráficos dos Resíduos versus Valores Ajustados.

Com esses resultados, podemos verificar a não robustez das inferências para os diferentes valores de β com distribuições diferentes da normal. Neste exemplo, $p(\theta_1, \theta_2 / \beta, \text{dados})$ muda sensivelmente quando β é modificado. Consequentemente, as inferências sobre $\theta = (\theta_1, \theta_2)$ diferem visivelmente dependendo do valor de β considerado.

Vamos considerar, agora, um conjunto com um grande número de observações para estudar esses efeitos.

x	y	x	y	x	y	x	y
0.02	34.270	0.10	102.062	0.50	167.210	0.90	179.737
0.02	34.110	0.10	99.078	0.50	167.206	0.90	181.163
0.02	32.599	0.10	98.189	0.50	167.542	0.90	179.861
0.02	33.437	0.10	101.207	0.50	164.489	0.90	179.777
0.02	35.750	0.10	100.653	0.50	165.003	0.90	178.428
0.03	46.723	0.20	134.599	0.60	169.596	1.00	182.776
0.03	46.571	0.20	132.361	0.60	171.189	1.00	181.645
0.03	44.622	0.20	132.626	0.60	174.103	1.00	180.739
0.03	43.782	0.20	133.870	0.60	172.383	1.00	183.187
0.03	45.914	0.20	133.704	0.60	172.205	1.00	180.187
0.05	66.983	0.30	149.736	0.70	175.264	1.10	181.962
0.05	66.609	0.30	151.165	0.70	177.211	1.10	182.215
0.05	66.908	0.30	148.798	0.70	174.824	1.10	180.849
0.05	65.297	0.30	149.622	0.70	176.777	1.10	184.925
0.05	68.273	0.30	150.720	0.70	175.081	1.10	181.458
0.08	86.710	0.40	159.678	0.80	176.881	1.20	183.966
0.08	88.646	0.40	161.269	0.80	178.116	1.20	183.187
0.08	90.262	0.40	158.685	0.80	178.270	1.20	185.569
0.08	89.390	0.40	160.477	0.80	180.060	1.20	183.824
0.08	88.737	0.40	158.823	0.80	176.884	1.20	184.407

Tabela 19. Dados gerados pelo modelo de Michaelis-Menten.

Considerando o conjunto de dados da tabela (19), com $n = 80$ observações, geradas pelo modelo de Michaelis-Menten (7.11), com $\theta_1 = 200$, $\theta_2 = 0.10$, $\sigma = 1$ e assumindo normalidade para os erros $\varepsilon_i \sim N(0,1)$, de (5.8), a densidade a posteriori conjunta para (θ_1, θ_2) , considerando β conhecido é dada por,

$$p(\theta_1, \theta_2 / \beta, \text{dados}) \propto [M(\theta_1, \theta_2)]^{-40(1+\beta)} \quad (7.13)$$

$$\text{onde } 0 < \theta_1, \theta_2 < \infty \text{ e } M(\theta_1, \theta_2) = \sum_{i=1}^{80} \left| y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right|^{2/(1+\beta)}.$$

A moda $(\hat{\theta}_1, \hat{\theta}_2)$ da densidade a posteriori conjunta (7.13) para cada valor de β considerado é dada na tabela (20). Para alguns valores de β , construímos o contorno da densidade a posteriori de (θ_1, θ_2) calculado na moda, através do software "Maple".

β	θ_1	θ_2
-0.90	200.060	0.099852
-0.85	200.077	0.099957
-0.75	200.085	0.100071
-0.65	200.064	0.100104
-0.50	200.012	0.100064
-0.40	199.975	0.100002
-0.30	199.937	0.099923
-0.25	199.918	0.099879
-0.20	199.900	0.099835
-0.10	199.862	0.099744
0.00	199.825	0.099654
0.10	199.787	0.099563
0.20	199.749	0.099474
0.25	199.731	0.099434
0.30	199.715	0.099393
0.40	199.685	0.099317
0.50	199.661	0.099245

Tabela 20. Moda de $p(\theta_1, \theta_2 / \beta, y)$ para vários valores de β .

Através dos contornos, figura (50) e das modas a posteriori apresentadas na tabela (20), observamos que as inferências sobre (θ_1, θ_2) não são muito afetadas pelos valores de β considerados, o que pode ser confirmado pelo gráfico das curvas ajustadas, figura (49), indicando a robustez das inferências obtidas com distribuições diferentes da distribuição normal.

Figura 49. Curvas ajustadas para os valores de β considerados na tabela (20): dados tabela (19).

(i) $\beta = -0.90$

(ii) $\beta = -0.60$

(iii) $\beta = -0.30$

(iv) $\beta = 0.00$

(v) $\beta = 0.30$

Figura 50. (i-v). Contornos da densidade a posteriori conjunta de (θ_1, θ_2) para vários valores de β .

Através dos contornos apresentados na figura (50), para este exemplo, consistindo de $n = 80$ observações, a forma e a posição dos contornos não é afetada sensivelmente quando consideramos diferentes valores para β . Em particular, os parâmetros θ_1 e θ_2 são positivamente correlacionados e a dispersão não muda significativamente quando consideramos diferentes valores para β . O mesmo acontece com as densidades a posteriori conjunta de (θ_1, θ_2) , (figura (51)-(i-iii)), que não mudam significativamente com a mudança de β .

(i) $\beta = -0.90$

(ii) $\beta = 0.00$

(iii) $\beta = 0.30$

Figura 51. (i-iii). Densidades a posteriori conjunta de (θ_1, θ_2) para vários valores de β .

Através da figura (52), observamos, claramente, uma boa linearidade dos gráficos normais probabilísticos, indicando que os resíduos são normalmente distribuídos, para qualquer um dos valores de β considerados.

(i) $\beta = -0.90$ (ii) $\beta = -0.6$ (iii) $\beta = -0.30$ (iv) $\beta = -0.20$ (v) $\beta = 0.00$ (vi) $\beta = 0.20$

Figura 52.(i-viii). Gráficos Normais Probabilísticos.

Através da figura (53), observamos os gráficos dos resíduos versus os valores ajustados, considerando vários valores para β , indicando uma boa aleatorização dos pontos e portanto, garantindo a homocedasticidade da variância, para qualquer um dos valores de β considerados.

Figura 53.(i-viii). Gráficos dos Resíduos versus Valores Ajustados.

Assim, considerando um conjunto de dados com um grande número ($n = 80$) de observações, as inferências são robustas para os diferentes valores de β considerados. Portanto, não existe sensibilidade das inferências obtidas e $p(\theta_1, \theta_2 / \beta, \text{dados})$ não muda sensivelmente quando β é modificado e consequentemente, as inferências sobre θ não diferem sensivelmente considerando qualquer valor de β .

7.4.2. Análise Bayesiana considerando β desconhecido

Considerando o conjunto de dados da tabela (1) e a densidade a priori uniforme (5.11), a densidade a posteriori marginal para β , aproximada pelo método de Laplace, é dada (ver 5.17) por,

$$p_u(\beta / \text{dados}) \propto [w(\beta)]^{12} [c(\beta)]^{-6(1+\beta)} \Gamma[6(1+\beta)] \frac{M^{-6(1+\beta)+1}}{[M_{11}M_{22} - (M_{12})^2]^{\frac{1}{2}}} \quad (7.14)$$

$$\text{onde } w(\beta) = \frac{\left\{ \Gamma\left[\frac{3}{2}(1+\beta)\right] \right\}^{\frac{1}{2}}}{(1+\beta) \left\{ \Gamma\left[\frac{1}{2}(1+\beta)\right] \right\}^{\frac{3}{2}}}; \quad c(\beta) = \left\{ \frac{\Gamma\left[\frac{3}{2}(1+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]} \right\}^{\frac{1}{1+\beta}}; \quad M, M_{11}, M_{22} \text{ e } M_{12} \text{ estão}$$

definidos em (5.17) e $-1 < \beta \leq 1$.

Na figura (54), temos o gráfico da densidade a posteriori (7.14). Dessa amostra com somente 12 observações, como era previsto, há pouca evidência sobre a forma da distribuição dos dados. Assim, através da análise Bayesiana, as conclusões dependem basicamente da densidade a priori assumida para β . Devemos estar cientes que, neste exemplo, estamos numa situação onde a suposição a priori sobre β deve ser dada com muita atenção e cuidado, pois os dados são poucos para fornecer, por si próprios, bastante informação. Se existirem arquivos de experimentos similares anteriores, devemos examiná-los para determinar se eles sugerem ou não a estrutura dos erros como sendo aproximadamente normal. Assim, neste exemplo, como a amostra é muito pequena, $n = 12$ observações, o uso da densidade a priori uniforme não nos fornece uma densidade a posteriori conveniente, pois não há informação suficiente para o modelo apropriado.

Figura 54. Dens. a Posteriori para β , considerando a priori uniforme: dados tabela (1).

Box e Tiao (1973), também consideram a priori informativa,

$$p(\beta) \propto (1 - \beta^2)^{a-1}, \quad -1 < \beta \leq 1 \quad (7.15)$$

onde o parâmetro a pode ser ajustado para representar um maior ou menor efeito do limite central. O resultado da variação de a é mostrado na figura (55), onde $p(\beta)$ e a correspondente densidade a posteriori de β ,

$$p(\beta / \text{dados}) \propto p_u(\beta / \text{dados}) p(\beta), \quad -1 < \beta \leq 1 \quad (7.16)$$

são representadas por uma curva tracejada e por uma curva contínua, respectivamente, para os valores de $a = 1, 3, 5, 7, 9, 11, 13, 15$ e 17 . Observamos, que com valores de a maiores que 13, assumindo informação razoável a priori, obtemos densidades a posteriori para β informativas.

(vii) $\alpha = 13$ (viii) $\alpha = 15$ (ix) $\alpha = 17$

Figura 55. (i-ix). Densidades a Priori e a Posteriori de β para várias escolhas do parâmetro α : dados da tabela (1).

Os gráficos da figura (55), mostram como o aumento da informação a priori sobre a normalidade tende a anular a informação proveniente da amostra de dados. Quando α tende ao infinito, $p(\beta / \text{dados})$ não se aproxima de uma função delta em $\beta = 0$ e a normalidade não pode ser assumida, ressaltando que o número de observações, para este exemplo, é pequeno.

Considerando o conjunto de dados da tabela (19), consistindo de 80 observações e a densidade a priori uniforme (5.11), a densidade a posteriori para β , obtida pelo método de Laplace, é dada por,

$$p_u(\beta / \text{dados}) \propto [w(\beta)]^{80} [c(\beta)]^{-40(1+\beta)} \Gamma[40(1+\beta)] \frac{M^{-40(1+\beta)+1}}{[M_{11}M_{22} - (M_{12})^2]^{\frac{1}{2}}} \quad (7.17)$$

onde $w(\beta)$ e $c(\beta)$ são definidos em (7.14) e M, M_{11}, M_{22} e M_{12} estão definidos em (5.17).

Observamos, que o uso da priori uniforme nos conduz a uma densidade a posteriori conveniente, pois há informação suficiente para o modelo apropriado. Como $p_u(\beta / \text{dados})$ é precisa, as inferências devem ser baseadas principalmente na informação contida nos dados, representada pela densidade $p_u(\beta / \text{dados})$ e não nas escolhas de $p(\beta)$. A moda da densidade (7.17) está razoavelmente próxima de zero e observamos que $\beta = 0$ está incluído em qualquer intervalo Bayesiano para β , indicando que a distribuição Normal é apropriada para os dados da tabela (19).

Figura 56. Densidade a Posteriori para β , considerando a priori uniforme: dados da tabela (19).

Considerando a priori informativa (7.15), o resultado da variação de a é mostrado na figura (57), onde $p(\beta)$ e a correspondente distribuição a posteriori de β , $p(\beta/dados) \propto p_u(\beta/dados)p(\beta)$, são representadas por curvas tracejada e contínua, respectivamente, para $a = 1, 2, 4, 6, 8$ e 10 . Para $a \geq 8$ a distribuição é próxima da priori. A figura (57) mostra que com o aumento da informação a priori sobre a Normalidade, não anulamos a informação proveniente dos dados e podemos assumir que a distribuição dos dados é, de fato, a distribuição Normal. Temos, portanto, conclusões similares com as obtidas utilizando a densidade a priori uniforme, isto é, a normalidade para este conjunto de dados ($\beta = 0$) é adequada.

(i) $a = 1$

(ii) $a = 2$

(iii) $a = 4$

Figura 57.(i-vi). Densidades a Priori e a Posteriori de β para várias escolhas do parâmetro α : dados tabela (19).

Assim, toda a inferência sobre (θ_1, θ_2) pode ser realizada a partir da densidade condicional, $p(\theta_j / \beta, \text{dados})$ considerando a densidade a priori (5.11). Para estes dados, através de suposições mais amplas, podemos confirmar a precisão da análise da teoria Normal.

O uso de métodos Bayesianos aproximados é uma alternativa apropriada para obter inferências precisas para os parâmetros do modelo de Michaelis-Menten (7.11), assumindo as suposições usuais, como vimos em exemplos anteriores. Também podemos usar os resultados Bayesianos aproximados para verificar as suposições usuais para os erros no modelo (7.11), considerando a densidade a posteriori marginal para β (ver (7.14) e (7.16)), assumindo a distribuição potência exponencial (5.2) para os erros, ou obter melhores inferências considerando o valor apropriado de β na distribuição (5.2).

Capítulo 8

CONCLUSÕES FINAIS E CONSIDERAÇÕES FUTURAS

Considerando o modelo não-linear de Michaelis-Menten, verificamos a importância de uma boa parametrização para a obtenção de resultados assintóticos precisos, independentemente do tamanho amostral ou do conjunto de dados considerado. Para isso, verificamos que é fundamental o uso de medidas de não-linearidade, de vícios e estudos de simulações que informam sobre a necessidade ou não de encontrar uma boa parametrização. Também destacamos o uso de algumas medidas recentemente introduzidas na literatura para diagnosticar a normalidade de funções de verossimilhança ou densidades a posteriori de interesse (Kass e Slate, 1992; Hills e Smith, 1993).

Também, verificamos que o uso de métodos Bayesianos aproximados pode ser uma alternativa muito promissora para analisar modelos não-lineares, considerando erros normais ou não-normais. Os resultados Bayesianos aproximados obtidos são de fácil implementação e com grande precisão e não requerem o uso de métodos computacionais sofisticados. Isso também foi verificado no problema de transformação dos dados. Os exemplos no final do trabalho, ilustram esses resultados.

Em relação às perspectivas de pesquisa futura, destacamos:

1. O uso de outros modelos não-lineares usados em aplicações.
2. O uso das medidas de Kass e Slate (1992) para encontrar parametrizações apropriadas em cada aplicação.
3. O uso de outras técnicas de integração para obter os resultados Bayesianos. Aqui poderíamos usar técnicas de Monte Carlo e "Gibbs Sampling".
4. Considerar técnicas Bayesianas para discriminar modelos não-lineares.
5. Considerações sobre planejamento de experimentos.

Apêndice A

Teoria Assintótica

Embora, frequentemente, usamos a expansão em série de Taylor para linearizar a teoria não-linear, assumimos que vários estimadores, tais como mínimos quadrados e estimadores de máxima verossimilhança são assintoticamente normais. A validade desses resultados requerem certas condições de regularidade, e nesta seção vamos apresentar algumas condições fundamentais para a teoria normal. A maioria da teoria é muito técnica, vamos apresentá-la sob o ponto de vista dos principais objetivos envolvidos, sem nos preocupar com as provas e detalhes técnicos.

A.1. Estimadores de Mínimos Quadrados

A.1.1. Existência dos Estimadores de Mínimos Quadrados

Consideramos o modelo não-linear usual,

$$y_i = f(x_i; \theta^*) + \varepsilon_i \quad (i = 1, 2, \dots, n), \quad (\text{A.1})$$

onde θ^* é o verdadeiro valor do vetor p -dimensional dos parâmetros θ e x_i é um vetor $(k \times 1)$ em $\mathbb{R}^k$. Assumindo que,

A(1). Os ε_i são i.i.d. com média zero e variância σ^2 ($\sigma^2 > 0$).

A(2). Para cada i , $f_i(\theta) = f(x_i; \theta)$ é uma função contínua de θ para $\theta \in \Theta$.

A(3). Θ é um subconjunto sólido de $\mathbb{R}^p$.

Qualquer valor de θ que minimiza,

$$S_n(\theta) = \sum_{i=1}^n [y_i - f_i(\theta)]^2$$

onde $\theta \in \Theta$ é chamado estimador de mínimos quadrados e será denotado por $\hat{\theta}_n$.

Dadas as suposições A(1) e A(3), Jennrich (1969) mostrou que o estimador de mínimos quadrados existe, isto é, existe uma função $\hat{\theta}_n = \hat{\theta}_n(y_1, y_2, \dots, y_n)$ das observações que minimiza $S_n(\theta)$. Agora, sob quais condições, $\hat{\theta}_n$ é:

- i) um estimador fracamente consistente de θ^* (isto é, $\text{plim } \hat{\theta}_n = \theta^*$, ou dado $\delta > 0$, $\lim \text{pr} [\|\hat{\theta}_n - \theta^*\| < \delta] = 1$), e
- ii) um estimador fortemente consistente ($\hat{\theta}_n \rightarrow \theta^*$, isto é, $\text{pr} [\hat{\theta}_n \rightarrow \theta^*] = 1$). Consistência forte implica em consistência fraca mas, o inverso não é válido.

A.1.2. Consistência

Amemiya (1983) fornece um esboço prestimoso dos ingredientes necessários para a consistência de $\hat{\theta}_n$. Ele indica quais os passos essenciais para provar que θ^* minimiza unicamente $\text{plim } n^{-1}S_n(\theta)$. Neste caso, se n for suficientemente grande tal que $n^{-1}S_n(\theta)$ seja próximo de $\text{plim } n^{-1}S_n(\theta)$, então $\hat{\theta}_n$ que minimiza $n^{-1}S_n(\theta)$ será próximo de θ^* que minimiza $\text{plim } n^{-1}S_n(\theta)$. Isto nos fornece a consistência fraca. Definindo,

$$\begin{aligned}
 n^{-1}S_n(\theta) &= n^{-1} \sum_i [y_i - f_i(\theta)]^2 \\
 &= n^{-1} \sum_i [y_i - f(\theta^*) + f_i(\theta^*) - f_i(\theta)]^2 \\
 &= n^{-1} \sum_i [\varepsilon_i + f_i(\theta^*) - f_i(\theta)]^2 \\
 &= n^{-1} \sum_i \varepsilon_i^2 + 2n^{-1} \sum_i \varepsilon_i [f_i(\theta^*) - f_i(\theta)] + n^{-1} \sum_i [f_i(\theta^*) - f_i(\theta)]^2 \\
 &= C_1 + C_2 + C_3
 \end{aligned}$$

Pela lei dos grandes números, $\text{plim } C_1 = \sigma^2$ e sujeito a certas condições em $n^{-1} \sum_i f_i(\theta^*) f_i(\theta)$ [ver suposição A(4a)] e $\text{plim } C_2 = 0$. Assim, $\text{plim } n^{-1}S_n(\theta)$ terá um único mínimo em θ^* se $\lim C_3$ também tiver um único mínimo em θ^* [ver a suposição A(4b)]. Dessa forma, a consistência fraca deve ser provada. A consistência forte de $\hat{\theta}_n$ é obtida usando os limites "almost sure" ao invés dos limites em probabilidade.

Com essa idéia em mente, definimos,

$$\begin{aligned}
 B_n(\theta, \theta_1) &= \sum_{i=1}^n f_i(\theta) f_i(\theta_1) \\
 D_n(\theta, \theta_1) &= \sum_{i=1}^n [f_i(\theta) f_i(\theta_1)]^2
 \end{aligned} \tag{A.1.2}$$

e adicionamos a seguinte suposição,

A(4). a) $n^{-1}B_n(\theta, \theta_1)$ converge uniformemente para θ, θ_1 em Θ para uma função $B_n(\theta, \theta_1)$ (que é contínua se (A2) e (A3) são verdadeiras). Implicando, pela expansão de (A.1.2), que $n^{-1}D_n(\theta, \theta_1)$ converge uniformemente para $D(\theta, \theta_1) = B(\theta, \theta) + B(\theta_1, \theta_1) - 2B(\theta, \theta_1)$.

b) Assumimos que $D_n(\theta, \theta^*) = 0$ se e somente se $\theta = \theta^*$ (isto é, $D_n(\theta, \theta^*)$ é definida positiva).

Dadas as suposições de *A(1)* a *A(3)*, Jennrich (1969) prova que *A(4)* é suficiente para $\hat{\theta}_n$ e $\hat{\sigma}_n^2 = S_n(\hat{\theta}_n) / n$ que são estimadores consistentes de θ^* e σ^2 , respectivamente.

A.1.3. Normalidade Assintótica

Assumindo, ao menos, consistência fraca do estimador de mínimos quadrados, vamos encontrar as condições nas quais $\hat{\theta}_n$ é assintoticamente normal.

Usando o teorema do valor médio e assumindo que as derivadas contínuas existem, com uma alta probabilidade para n grande, temos,

$$\begin{aligned} 0 &= \frac{\partial S_n(\hat{\theta}_n)}{\partial \theta} \\ &= \frac{\partial S_n(\theta^*)}{\partial \theta} + \frac{\partial^2 S_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta^*) \end{aligned} \quad (\text{A.1.3.1})$$

onde $\partial S_n(\theta^*) / \partial \theta$ representa $\partial S_n(\theta) / \partial \theta$ calculado em $\theta = \theta^*$, e $\tilde{\theta}_n$ está entre $\hat{\theta}_n$ e θ^* . Então, de (A.1.3.1), temos,

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = - \left[\frac{1}{n} \frac{\partial^2 S_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} \right]^{-1} \frac{1}{\sqrt{n}} \frac{\partial S_n(\theta^*)}{\partial \theta} \quad (\text{A.1.3.2})$$

precisamos provar que,

i) $n^{-1/2} \partial S_n(\theta^*) / \partial \theta$ é assintoticamente normal, e

ii) $n^{-1} \partial^2 S_n(\tilde{\theta}_n) / \partial \theta \partial \theta'$ converge em probabilidade para uma matriz não singular.

i) Inicialmente, temos,

$$\frac{\partial S_n(\theta)}{\partial \theta} = -2 \sum_{i=1}^n [y_i - f_i(\theta)] \frac{\partial f_i(\theta)}{\partial \theta} \quad (\text{A.1.3.3})$$

tal que

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial S_n(\theta^*)}{\partial \theta} &= -\frac{2}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \frac{\partial f_i(\theta^*)}{\partial \theta} \\ &= -\frac{2}{\sqrt{n}} \mathbf{F}'(\theta^*) \boldsymbol{\varepsilon} \end{aligned} \quad (\text{A.1.3.4})$$

onde $\mathbf{F}(\theta) = \left[\left(\partial f_i(\theta) / \partial \theta_r \right) \right]$. Desde que (A.1.3.4) represente uma média ponderada da variável i.i.d. ε_i , podemos aplicar o teorema do limite central e assim obter,

$$\frac{1}{\sqrt{n}} \frac{\partial S_n(\theta^*)}{\partial \theta} \sim N_p(0, 4\sigma^2 \Omega), \text{ assintoticamente,} \quad (\text{A.1.3.5})$$

contanto que,

$$\begin{aligned} \Omega &= \lim \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial f_i(\theta^*)}{\partial \theta} \frac{\partial f_i(\theta^*)}{\partial \theta'} \right\} \\ &= \lim \frac{1}{n} \mathbf{F}'(\theta^*) \mathbf{F}(\theta^*) \end{aligned} \quad (\text{A.1.3.6})$$

exista e seja não singular.

ii) Diferenciando (A.1.3.3), obtemos,

$$\begin{aligned} \frac{1}{n} \frac{\partial^2 S_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} &= \frac{2}{n} \sum_i \frac{\partial f_i(\tilde{\theta}_n)}{\partial \theta} \frac{\partial f_i(\tilde{\theta}_n)}{\partial \theta'} \\ &\quad - \frac{2}{n} \sum_i [f_i(\theta^*) - f_i(\tilde{\theta}_n)] \frac{\partial^2 f_i(\tilde{\theta}_n)}{\partial \theta \partial \theta'} - \frac{2}{n} \sum_i \varepsilon_i \frac{\partial^2 f_i(\tilde{\theta}_n)}{\partial \theta \partial \theta'} \\ &= G_1 - G_2 - G_3 \end{aligned} \quad (\text{A.1.3.7})$$

Se $\hat{\theta}_n$ for consistente, então $\tilde{\theta}_n \rightarrow \theta^*$ com probabilidade um. Portanto, com probabilidade aproximadamente um quando $n \rightarrow \infty$, $\tilde{\theta}_n$ estará em Θ . Agora, se $h_n(\theta)$ for uma sequência qualquer de funções contínuas e $h_n(\theta)$ converge certamente para uma função não estocástica $h(\theta)$ uniformemente em θ , então,

$$\text{plim } h_n(\tilde{\theta}_n) = h(\text{plim } \tilde{\theta}_n) = h(\theta^*).$$

Antes de provar *ii*), vamos aplicar o resultado acima para cada elemento da equação (A.1.3.7) e mostraremos que cada um dos três termos do lado direito dessa equação converge para uma função não estocástica uniformemente em θ . Condições suficientes para isso são dadas por,

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial f_i(\theta)}{\partial \theta_r} \frac{\partial f_i(\theta)}{\partial \theta_s} \quad \text{e} \quad \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial^2 f_i(\theta)}{\partial \theta_r \partial \theta_s} \right]^2$$

que convergem uniformemente para $\theta \in \Theta$ e $r, s=1, 2, \dots, p$. Estas condições implicam que $\text{plim } G_\alpha = 0$ ($\alpha=2, 3$) e usando (A.1.3.6), observamos que (A.1.3.7) é dado por,

$$\text{plim } \frac{1}{n} \frac{\partial^2 S_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} = 2\Omega \quad (\text{A.1.3.8})$$

Portanto, de (A.1.3.2), (A.1.3.4) e (A.1.3.8) temos,

$$\sqrt{n}(\hat{\theta} - \theta^*) \approx -(2\Omega)^{-1} \frac{1}{\sqrt{n}} \frac{\partial S_n(\theta^*)}{\partial \theta} \quad (\text{A.1.3.9})$$

$$\rightarrow N_p(0, \sigma^2 \Omega^{-1}) \quad [\text{por (A.1.3.5)}] \quad (\text{A.1.3.10})$$

Mas *i*) e *ii*) implicam em (A.1.3.10). De (A.1.3.2) e (A.1.3.4), (A.1.3.9) torna-se,

$$\hat{\theta}_n - \theta^* \approx [F'(\theta^*) F(\theta^*)]^{-1} F'(\theta^*) \varepsilon \quad (\text{A.1.3.11})$$

Com o propósito de resumir a discussão acima, vamos listar as suposições necessárias para (A.1.3.10) ser verdadeira.

A(5). θ^* é um ponto interior de Θ .

A(6). As primeiras e segundas derivadas $\partial f_i(\theta) / \partial \theta_r$ e $\partial^2 f_i(\theta) / \partial \theta_r \partial \theta_s$ ($r, s=1, 2, \dots, p$), existem e são contínuas para todo $\theta \in \Theta$.

A(7). $(1/n) \sum_{i=1}^n (\partial f_i(\theta) / \partial \theta) (\partial f_i(\theta) / \partial \theta)' [= n^{-1} F'(\theta) F(\theta)]$ converge para uma matriz $\Omega(\theta)$ uniformemente em θ para $\theta \in \Theta$.

A(8). $(1/n) \sum_{i=1}^n [\partial^2 f_i(\theta) / \partial \theta_r \partial \theta_s]^2$ converge uniformemente em θ para $\theta \in \Theta$ ($r, s=1, 2, \dots, p$).

A(9). $\Omega = \Omega(\theta^*)$ é não singular.

Se as suposições de A(1) a A(9) são satisfeitas, então,

$$\sqrt{n}(\hat{\theta} - \theta^*) \sim N_p(0, \sigma^2 \Omega^{-1}), \text{ assintoticamente.} \quad (\text{A.1.3.12})$$

E $n^{-1} \mathbf{F}'(\hat{\theta}_n) \mathbf{F}(\hat{\theta}_n)$ é um estimador consistente de Ω . De A(7) e da consistência forte de $\hat{\theta}_n$ [por A(1) e A(4)]. Finalmente, usando (A.1.3.6) temos, para n grande,

$$\hat{\theta} - \theta^* \sim N_p(0, \sigma^2 [\mathbf{F}'(\theta^*) \mathbf{F}(\theta^*)]^{-1}) \text{ aproximadamente.} \quad (\text{A.1.3.13})$$

Apêndice B

Decomposição QR de uma Matriz

Para calcular a decomposição QR de uma matriz $X(n \times p)$, usamos as transformações Householder (Householder, 1958), isto é, uma generalização de reflexões no plano. Essas matrizes $(n \times n)$ são da forma,

$$H_u = I - 2uu'$$

onde I $(n \times n)$ é uma matriz identidade e u é um vetor n -dimensional com componentes unitários $u' = (1, 1, \dots, 1)'$, assim, $\|u\| = \sqrt{u'u} = 1$. A matriz H_u é simétrica e ortogonal, pois,

$$H_u' = I' - 2uu' = H_u$$

e

$$H_u' H_u = I - 4uu' + 4uu'uu' = I$$

Multiplicando-se um vetor y por H_u , tal que,

$$H_u y = y - 2uu'y$$

corresponde à refletir y sobre uma linha passando pela origem e perpendicular à u , conforme a Figura (A.1). Escolhendo,

$$u = \frac{y - \|y\|e_1}{\|(y - \|y\|e_1)\|} \quad (A.1)$$

ou

$$u = \frac{y + \|y\|e_1}{\|(y + \|y\|e_1)\|} \quad (A.2)$$

onde $e_1 = (1, 0, \dots, 0)'$, a transformação Householder pode ser feita tomando-se o vetor y no eixo x , que corresponde, no novo sistema de coordenadas (rotação), ao vetor $H_u y$ que tem como coordenadas $(\pm\|y\|, 0, \dots, 0)'$.

Figura A.1 - Uma transformação Householder mostrando a reflexão do vetor y sobre a linha perpendicular à u .

Exemplo A.1 - Para efetuar a decomposição QR da matriz,

$$X = \begin{bmatrix} 1 & 1.26 \\ 1 & 1.82 \\ 1 & 2.22 \end{bmatrix}$$

escolhemos uma transformação H_{u_1} tomando-se a primeira coluna x_1 de X e usando (A.1) obtemos,

$$u_1 = \frac{(1,1,1)' - \sqrt{3}(1,0,0)'}{1.5925} = \frac{(-0.7321, 1, 1)'}{1.5925}$$

$$\begin{aligned} H_{u_1} &= I - 2u_1u_1' \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} -0.45969 \\ 0.62794 \\ 0.62794 \end{bmatrix} \begin{bmatrix} -0.45969 & 0.62794 & 0.62794 \end{bmatrix} \\ &= \begin{bmatrix} 0.57737 & 0.57731 & 0.57731 \\ 0.57731 & 0.21138 & -0.78862 \\ 0.57731 & -0.78862 & 0.21138 \end{bmatrix} \end{aligned}$$

Assim,

$$\mathbf{X}_1 = \mathbf{H}_{u_1} \mathbf{X} = \begin{bmatrix} 1.7321 & 3.0600 \\ 0 & -0.6388 \\ 0 & -0.2388 \end{bmatrix}$$

Agora calculamos a segunda rotação, ortogonal à primeira, escolhendo $\mathbf{u}_2$ tal que $\mathbf{H}_{u_2}$ tenha valores que zerem os elementos abaixo da diagonal da segunda coluna de $\mathbf{X}_1$. Para assegurar que não mudemos a primeira coluna, devemos então considerar o primeiro elemento de $\mathbf{u}_2$ igual a zero. Assim, escolhemos $\mathbf{u}_2$ tal que,

$$\mathbf{u}_2 = \frac{(0, -0.6388, -0.2388)' - 0.6820(0, 1, 0)'}{1.3422}$$

$$\begin{aligned} \mathbf{H}_{u_2} &= \mathbf{I} - 2\mathbf{u}_2\mathbf{u}_2' \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.9367 & -0.3502 \\ 0 & -0.3502 & 0.9367 \end{bmatrix} \end{aligned}$$

Portanto, temos,

$$\mathbf{R} = \mathbf{H}_{u_1} \mathbf{H}_{u_2} \mathbf{X} = \begin{bmatrix} 1.7321 & 3.0600 \\ 0 & 0.6820 \\ 0 & 0 \end{bmatrix}$$

A matriz $\mathbf{R}$ é calculada ao mesmo tempo que são obtidos os vetores $\mathbf{u}_1, \dots, \mathbf{u}_p$ da reflexão de Householder. A matriz $\mathbf{Q}$ correspondente é determinada à partir da relação $\mathbf{R} = \mathbf{Q}'\mathbf{X}$, isto é, $\mathbf{Q}' = \mathbf{H}_{u_p} \dots \mathbf{H}_{u_1}$. Do fato das transformações Householder serem simétricas, temos,

$$\mathbf{Q} = \mathbf{H}_{u_1}' \dots \mathbf{H}_{u_p}' = \mathbf{H}_{u_1} \dots \mathbf{H}_{u_p}$$

Observamos que a matriz $\mathbf{Q}$ é calculada de maneira explícita.

Para maiores informações sobre a decomposição QR ver, por exemplo, Dongarra et al.(1979) e Stewart (1973).

Apêndice C

Estudo de Simulação

Um estudo de simulação pode nos auxiliar na verificação do comportamento não-normal dos estimadores e possivelmente sugerir reparametrizações adequadas. O objetivo desse estudo é verificar se de fato as propriedades assintóticas dos estimadores de mínimos quadrados constituem uma boa aproximação para o problema, isto é, se de fato o nosso particular modelo tem comportamento próximo do linear. Num estudo de simulação, devemos conhecer o modelo não linear e a distribuição dos erros aleatórios ε_t ,

$$y_t = f(x_t; \theta) + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (\text{C.1})$$

onde f tem uma forma conhecida e $\varepsilon_t \sim N(0, \sigma^2)$ e independentes.

Os valores $\varepsilon_1, \dots, \varepsilon_n$ são gerados pseudo-aleatoriamente através de uma sub-rotina de computador, assumindo que ε_t são i.i.d. normais com média zero e variância σ^2 . Além disso, é fixado um valor para θ e um valor para σ^2 , para os quais usualmente utilizamos as estimativas de mínimos quadrados $\hat{\theta}$ e $\hat{\sigma}^2$, obtidas do conjunto original de dados e assim obtemos um conjunto de n observações $y_1, y_2, \dots, y_n$, onde $y_t = f(x_t; \theta) + \varepsilon_t$.

Simulamos N desses conjuntos (de n observações) e para cada conjunto ajustamos o modelo por mínimos quadrados da maneira usual, por exemplo, utilizando o método iterativo de Gauss Newton, onde podemos utilizar como valor inicial, em cada conjunto, a estimativa $\hat{\theta}$ obtida do conjunto original de dados. Se para cada um dos N conjuntos de dados, a convergência ocorrer, teremos N vetores de estimativas do tipo,

$$\hat{\theta}_k = (\hat{\theta}_{k1}, \hat{\theta}_{k2}, \dots, \hat{\theta}_{kp}) , \quad k = 1, 2, \dots, N \quad (\text{C.2})$$

Ratkowsky (1983) considera bastante satisfatório $N = 1000$ amostras para estudar as propriedades dos estimadores nos modelos de regressão não-linear. Assim, devemos examinar a distribuição conjunta dessas estimativas. Mas sabemos que a análise de uma distribuição conjunta é relativamente difícil, mesmo para o caso mais simples onde $p = 2$, em que devemos examinar se $\hat{\theta}$ tem uma distribuição normal bivariada. Ratkowsky (1983) usa, como uma aproximação, os resultados das análises das distribuições marginais de $\hat{\theta}$, no

intuito de que, através desse procedimento possamos avaliar aproximadamente o comportamento conjunto dos parâmetros.

Podemos analisar as distribuições marginais com base nas N estimativas dos parâmetros correspondentes, através de testes e através de uma análise descritiva, como por exemplo, o cálculo dos momentos amostrais e a construção de histogramas (nesse caso $\hat{\theta}$ deve ser padronizado).

A seguir, apresentaremos algumas medidas importantes para uma análise descritiva das distribuições, a saber,

$$\text{Média amostral:} \quad m1 = \bar{\hat{\theta}} = \sum_{j=1}^N \frac{\hat{\theta}_{ij}}{N}; \quad (\text{C.3})$$

$$\text{Variância amostral:} \quad m2 = s_i^2 = \frac{1}{N} \sum_{j=1}^N (\hat{\theta}_{ij} - \bar{\hat{\theta}})^2; \quad (\text{C.4})$$

$$\text{Coeficiente de "Skewness":} \quad g_{1i} = \frac{m_{3i}}{(m2)^{3/2}}; \quad (\text{C.5})$$

$$\text{Coeficiente de "Curtose":} \quad g_{2i} = \frac{m_{4i}}{(m2)^2}. \quad (\text{C.6})$$

onde:

$$m_{3i} = \frac{1}{N} \sum_{j=1}^N (\hat{\theta}_{ij} - \bar{\hat{\theta}})^3, \quad i=1, \dots, p \quad (\text{C.7})$$

e

$$m_{4i} = \frac{1}{N} \sum_{j=1}^N (\hat{\theta}_{ij} - \bar{\hat{\theta}})^4, \quad i=1, \dots, p \quad (\text{C.8})$$

O coeficiente de Skewness (C.5) mede a assimetria da distribuição dos $\hat{\theta}$ e o coeficiente de Curtose (C.6) mede o grau de "achatamento" da distribuição. As fórmulas em (C.7) e (C.8) representam, respectivamente, o terceiro e quarto momentos amostrais em torno da média.

Uma medida importante em nosso estudo é a estimativa do vício de $\hat{\theta}$ que é dada pela diferença entre a média amostral (C.3) e o verdadeiro valor do parâmetro (o valor de θ_i

usado para gerar os dados, que são, usualmente as estimativas de mínimos quadrados). O vício relativo é dado pela fórmula,

$$\frac{\bar{\hat{\theta}} - \theta_i}{\theta_i} \times 100$$

onde $\bar{\hat{\theta}}$ é a média amostral e θ_i o verdadeiro valor do parâmetro.

Podemos também calcular a variância assintótica de $\hat{\theta}$ dada por,

$$\sigma_{*i}^2 = \hat{\sigma}^2 \left[\sum_{t=1}^n \frac{\partial f(x_t; \hat{\theta})}{\partial \theta_i} \frac{\partial f(x_t; \hat{\theta})}{\partial \theta_j} \right]^{-1} \quad (C.9)$$

onde θ_i e θ_j são os elementos do vetor de parâmetros θ ($i, j = 1, 2, \dots, p$) e $\hat{\sigma}^2$ é a variância residual do conjunto original de dados.

A fórmula,

$$\frac{s_i^2 - \sigma_{*i}^2}{\sigma_{*i}^2} \times 100$$

onde s_i^2 é a variância amostral dada em (C.4), representa a porcentagem na qual a variância amostral excede a variância assintótica (C.9). Essa também é uma medida importante para a análise da distribuição de $\hat{\theta}$. Além do cálculo dessas medidas, também podemos realizar testes de hipóteses para verificar se a variância é mínima, o vício é zero e se os coeficientes de Curtose e Skewness são respectivamente 0 e 3 (os valores esperados sob amostragem de uma distribuição normal).

Para verificar se a suposição de variância mínima é válida, podemos comparar a variância amostral com a variância assintótica obtida, usando um teste Qui-Quadrado,

$$\chi^2 = \frac{(N-1)s_i^2}{\sigma_{*i}^2}$$

onde χ^2 tem uma distribuição Qui-Quadrado com $n-1$ graus de liberdade. Como nos estudos de simulação N é grande, podemos usar a aproximação normal padrão,

$$Z = \sqrt{2\chi^2} - \sqrt{2(m)-1}$$

onde m corresponde ao número de graus de liberdade da distribuição χ^2 .

Para verificar se o vício é zero, podemos utilizar a estatística,

$$Z = \frac{\bar{\theta} - \theta_i}{\sqrt{\sigma_{\theta_i}^2 / N}}$$

onde consideramos Z normalmente distribuída com média zero e variância um.

Para testar se o coeficiente de Skewness é 0 e o coeficiente de Curtose é 3 utilizamos o fato de que $g_1 \sim N(0; 6/N)$ e $g_2 \sim N(0; 24/N)$, (Snedecor e Cochran, 1980), então para $N = 1000$, os valores críticos bicaudais 5% e 1% para g_1 são $\pm 0,152$ e $\pm 0,200$, respectivamente e para g_2 os valores críticos são $\pm 0,304$ e $\pm 0,399$, respectivamente.

Se os testes forem todos significativos então existem evidências de que o estimador de mínimos quadrados não terá comportamento próximo do linear. Entretanto, em alguns casos podemos reduzir essa não-linearidade através de uma reparametrização. Quando temos um conjunto de resultados de simulação é muito simples examinar o efeito de reparametrizações, pois o estimador de mínimos quadrados de uma função do parâmetro é simplesmente a função do estimador de mínimos quadrados do parâmetro, porém a dificuldade que surge é na escolha da reparametrização. As medidas de não-linearidade apresentadas no capítulo 1, apesar de serem muito úteis para detectar o comportamento não-linear, não fornecem nenhuma "pista" nesse sentido; nesses casos o uso de simulação poderá indicar ou sugerir possíveis reparametrizações para o modelo.

Para maiores informações sobre o estudo de simulação ver por exemplo, Ratkowsky (1983).

Apêndice D

Alguns Métodos de Resolução de Integrais em Inferência Bayesiana

Usualmente temos interesse em calcular densidades preditivas e densidades a posteriori marginais, bem como suas médias e intervalos de credibilidade Bayesianos, que não envolvam densidades a priori conjugadas, assim, a resolução das integrais torna-se demasiadamente complicada, pois não podemos resolvê-las analiticamente, necessitando de um método numérico ou de aproximação para as integrais.

Nesta seção, vamos apresentar alguns métodos usuais para a integração em inferência Bayesiana, pois muitas vezes a resolução de integrais não apresentam soluções analíticas, exigindo o uso de métodos de aproximação de integrais (ver por exemplo, Lindley, 1980; Tierney e Kadane, (1986); Tierney, Kass e Kadane, (1989)) ou métodos numéricos (ver, por exemplo, Naylor e Smith, (1982)).

D.1. Aproximação de Tierney e Kadane

Muitas vezes, a resolução de integrais não apresentam soluções analíticas, exigindo o uso de métodos de aproximação. Neste sentido, um dos métodos mais utilizados em inferência Bayesiana é o *método de Laplace* para aproximação de momentos a posteriori introduzido por Tierney e Kadane (1986).

O método de Laplace para aproximação de integrais é usado para resolver integrais da seguinte forma: supondo que desejamos resolver a integral em θ , dada por,

$$I = \int_{\Theta} r(\theta) d\theta \quad (D.1)$$

Uma aproximação para a integral em (D.1), pelo método de Laplace, é dada por,

$$\int_{\Theta} r(\theta) d\theta = \int_{\Theta} f(\theta) e^{-nh(\theta)} d\theta, \quad (D.2)$$

onde $h(\theta)$ é uma função unimodal com máximo em $\hat{\theta}$ e satisfaz algumas condições de regularidade. O método de Laplace assume um desenvolvimento de h e f em série de *Taylor*, em torno de $\hat{\theta}$, para aproximar integrais da forma (D.2). Para maiores informações ver Tierney e Kadane, (1986).

No caso em que θ é uniparamétrico, a integral (D.1) é aproximada por,

$$I \approx \left(\frac{2\pi}{n}\right)^{1/2} f(\hat{\theta}) [nh''(\hat{\theta})]^{-1/2} e^{-nh(\hat{\theta})}. \quad (D.3)$$

Com $\theta \in \mathbb{R}^p$, caso multiparamétrico, a aproximação para a integral em (D.1) é dada por,

$$I \approx (2\pi)^{p/2} \left\{ \det[nD^2h(\hat{\theta})] \right\}^{-1/2} f(\hat{\theta}) e^{-nh(\hat{\theta})}. \quad (D.4)$$

onde p representa o número de parâmetros e $D^2h(\hat{\theta})$ é a matriz Hessiana de h calculada em $\hat{\theta}$, definida por,

$$D^2h(\theta) = \begin{bmatrix} \frac{\partial^2 h}{\partial \theta_1^2} & \frac{\partial^2 h}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 h}{\partial \theta_1 \partial \theta_p} \\ & \frac{\partial^2 h}{\partial \theta_2^2} & \cdots & \frac{\partial^2 h}{\partial \theta_2 \partial \theta_p} \\ & & \ddots & \vdots \\ \text{simétrica} & & & \frac{\partial^2 h}{\partial \theta_p^2} \end{bmatrix}.$$

Em aplicações Bayesianas, com θ uniparamétrico, supondo que desejamos calcular a esperança a posteriori de uma função $r(\theta)$, monótona crescente, dada por,

$$E_0[r(\theta) / D] = \frac{\int_{\Theta} r(\theta) p(y / \theta) p(\theta) d\theta}{\int_{\Theta} p(y / \theta) p(\theta) d\theta}, \quad (D.5)$$

onde $p(\theta)$ é a densidade a priori de θ e $p(y/\theta)$ ou $L(\theta)$ representa a função de verossimilhança. De (D.2) com $f(\theta) = 1$, podemos reescrever (D.5) na forma,

$$E_{\theta}[r(\theta) / D] = \frac{\int_{\Theta} e^{-nh_*(\theta)} d\theta}{\int_{\Theta} e^{-nh(\theta)} d\theta}, \quad (\text{D.6})$$

onde $-nh_*(\theta) = \ln r(\theta) + \ln p(y / \theta) + \ln p(\theta)$ e $-nh(\theta) = \ln p(y / \theta) + \ln p(\theta)$. Assim, aplicando o método de Laplace para o numerador e o denominador de (D.6), obtemos uma aproximação para a esperança a posteriori (D.5) dada por,

$$E_{\theta}[r(\theta) / D] \approx \left[\frac{nh'_*(\hat{\theta}_*)}{nh''(\hat{\theta})} \right]^{-1/2} \exp\{-[nh_*(\hat{\theta}_*) - nh(\hat{\theta})]\}, \quad (\text{D.7})$$

onde $\hat{\theta}$ maximiza $-nh(\theta)$ e $\hat{\theta}_*$ maximiza $-nh_*(\theta)$.

Podemos considerar outros valores para $f(\theta)$ nas integrais (D.2), na tentativa de obter aproximações mais precisas e simplificação dos cálculos. Devemos ressaltar, que o procedimento de Laplace é sensível às parametrizações, ou seja, uma boa reparametrização fornece aproximações mais precisas, especialmente para amostras pequenas, pois para grandes amostras conseguimos boa aproximação para (D.7). Mais detalhes sobre a precisão do método de Laplace descrito, podem ser encontrados em Tierney e Kadane, (1986) e para a obtenção de boas reparametrizações, ver, por exemplo, Achcar e Smith, (1990); ou Kass e Slate (1992).

D.2. Métodos Numéricos

O uso de métodos numéricos é uma outra estratégia explorada por muitos estatísticos Bayesianos (ver, por exemplo, Naylor e Smith, (1982)). Usualmente, estes métodos são apropriados para problemas que apresentem um pequeno número de parâmetros.

Nesta seção, vamos apresentar um breve sumário de alguns dos procedimentos numéricos disponíveis para a resolução de integrais em estatística Bayesiana.

D.2.1. Regra de Simpson

Um dos métodos utilizados em estatística Bayesiana é a regra de Simpson, que para o caso uniparamétrico é definida por,

$$\int_a^b f(\theta) d\theta = \frac{h}{3} (f_0 + 4f_1 + 2f_2 + 4f_3 + \dots + 2f_{2n-2} + 4f_{2n-1} + f_{2n}) \quad (\text{D.8})$$

onde $f_r = f(\theta_r)$, isto é,

$$\begin{cases} f_0 = f(\theta_0) \\ f_1 = f(\theta_0 + h) = f(\theta_1) \\ f_2 = f(\theta_1 + h) = f(\theta_2) \\ \vdots \\ f_n = f(\theta_{n-1} + h) = f(\theta_n) \end{cases}$$

e o intervalo finito $[a, b]$ é dividido em $2n$ (número par) de subintervalos, cada um de comprimento h , tal que $b - a = 2nh$, $[(b - a)/h = 2n]$. O erro da fórmula (D.8) é definido por,

$$R_n = -\frac{1}{180} (b - a) h^4 f^{(4)}(\xi), \quad \text{onde } a < \xi < b.$$

A regra de Simpson é apropriada quando o intervalo $[a, b]$ é finito e $f(\theta)$ é diferenciável, mas isso pode não ser possível para muitas aplicações Bayesianas e às vezes, $f(\theta)$ pode ser muito complicada.

D.2.2. Regra Gaussiana

Naylor e Smith (1982) usam procedimentos de Quadratura Gaussiana de integração numérica para resolver integrais Bayesianas. Para usar a regra Gaussiana, escolhemos uma função integrável não negativa $\omega(\theta)$, tal que,

$$\int_a^b f(\theta) d\theta = \int_a^b r_G(\theta) \omega(\theta) d\theta \quad (\text{D.9})$$

onde $r_G(\theta) = f(\theta) / \omega(\theta)$ é uma função bem comportada.

Um possível procedimento é dado por polinômios ortogonais (ver por exemplo, Conte e De Boor, (1965)) para encontrar métodos eficientes para aproximar (D.5).

Um resultado importante em integração numérica é que se $P_{2n-1}(\theta)$ representa um polinômio qualquer de grau $(2n-1)$ em θ , definido em $[a, b]$ e $\{f_n(\theta)\}$ representa uma

sequência de polinômios ortogonais em $[a, b]$ relativo a uma função de peso positivo $\omega(\theta)$, tal que $f_n(\theta)$ seja um polinômio positivo, então,

$$\int_a^b \omega(\theta) P_{2n-1}(\theta) d\theta = \sum_{k=1}^n \alpha_k P_{2n-1}(\theta_k) \quad (\text{D.10})$$

onde α_k representam os coeficientes e θ_k representam as raízes do polinômio $f_n(\theta) = 0$. Os coeficientes são dados por,

$$\alpha_k = \int_a^b \frac{\omega(\theta) f_n(\theta) d\theta}{(\theta - \theta_k) f_n(\theta_k)}$$

Na prática, podemos usar diferentes sistemas de polinômios ortogonais. Uma possibilidade, explorada por Naylor e Smith (1982) é dada pela classe de polinômios ortogonais relativo à função de peso $\omega(\theta) = e^{-\theta^2}$, no intervalo $[a, b] = [-\infty, \infty]$.

Assim, temos a aproximação,

$$\int_{-\infty}^{\infty} e^{-\theta^2} r_G(\theta) d\theta \approx \sum_{k=1}^n \alpha_k r_G(\theta_k) + R_n \quad (\text{D.11})$$

onde $\theta_1, \theta_2, \dots, \theta_n$ são as raízes da equação polinomial de Hermite $H_n(\theta) = 0$, e

$$\alpha_k = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(\theta_k)]^2} \quad (\text{D.12})$$

Os valores das raízes θ_k e os coeficientes α_k são dados em tabelas (Abramowitz e Stegun, (1964), p. 924).

Apêndice E

E.1. Integral Gama Invertida

Para $a > 0$, $p > 0$,

$$\int_0^{\infty} x^{-(p+1)} e^{-ax^{-2}} dx = \frac{1}{2} a^{-p/2} \Gamma(p/2) \quad (\text{E.1.1})$$

(ver por exemplo, Abramowitz e Stegun, (1968)).

E.2. Expressões utilizadas para a obtenção de densidades a posteriori marginais de θ (ver Capítulo 4) :

E.2.1. Densidade a posteriori marginal de θ_1 :

$$S(\theta_1, \theta_2) = \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \quad (\text{E.2.1.1})$$

$$S_2(\theta_1, \theta_2) = 2\theta_1 \left\{ \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - \theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right\} \quad (\text{E.2.1.2})$$

$$S_{22}(\theta_1, \theta_2) = -4\theta_1 \left\{ \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^3} \right\} + 6\theta_1^2 \left\{ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right\} \quad (\text{E.2.1.3})$$

$$\begin{aligned} g'(\theta_2) = & -2 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right) - 4 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^5} \right) + \\ & + 6 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right) \end{aligned} \quad (\text{E.2.1.4})$$

$$g''(\theta_2) = -12 \left[\left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right) \right]^2 - 8 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^5} \right) + \\ + 20 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^6} \right) \quad (\text{E.2.1.5})$$

E.2.2. Densidade a posteriori marginal para θ_2 :

$$S_1(\theta_1, \theta_2) = 2\theta_1 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} \right) - 2 \left(\sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right) \quad (\text{E.2.2.1})$$

$$S_{11}(\theta_1, \theta_2) = 2 \left(\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} \right) \quad (\text{E.2.2.2})$$

E.3. Expressões utilizadas para a obtenção de momentos a posteriori para θ

E.3.1. Média a posteriori para θ_1 :

O determinante da matriz Hessiana do numerador de (4.26) é dado por:

$$\det(nD^2 h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)) = A^*(\hat{\theta}_1, \hat{\theta}_2) B^*(\hat{\theta}_1, \hat{\theta}_2) - [C^*(\hat{\theta}_1, \hat{\theta}_2)]^2 \quad (\text{E.3.1.1})$$

onde

$$A^*(\theta_1, \theta_2) = \frac{2}{\theta_1^2} + n \left\{ \frac{\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2}}{\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2} \right\} + \left\{ \frac{4\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - 4 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)}}{\theta_1 \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2} \right\},$$

$$B^*(\theta_1, \theta_2) = -\frac{g''(\theta_2)}{2g(\theta_2)} + \frac{1}{2} \left(\frac{g'(\theta_2)}{g(\theta_2)} \right)^2 - n\theta_1 \left\{ \frac{\left[2 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^3} - 3\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right]}{S(\theta_1, \theta_2)} \right\} -$$

$$-n\theta_1 \left\{ \frac{\left[\sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - \theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

onde $S(\theta_1, \theta_2)$; $g'(\theta_2)$ e $g''(\theta_2)$ estão definidos em (E.2.1.1); (E.2.1.4) e (E.2.1.5), respectivamente.

$$C^*(\theta_1, \theta_2) = \left\{ \frac{\left[-2n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} + n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} \right]}{S(\theta_1, \theta_2)} \right\} -$$

$$- \left\{ \frac{\left[n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

O determinante da matriz Hessiana do denominador de (4.26) é dado por:

$$\det(nD^2 h(\hat{\theta}_1, \hat{\theta}_2)) = A(\hat{\theta}_1, \hat{\theta}_2)B(\hat{\theta}_1, \hat{\theta}_2) - [C(\hat{\theta}_1, \hat{\theta}_2)]^2 \quad (\text{E.3.1.2})$$

onde:

$$A(\theta_1, \theta_2) = \frac{1}{\theta_1^2} + n \left\{ \frac{\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2}}{S(\theta_1, \theta_2)} \right\} - \left\{ \frac{2 \left[\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right]}{\theta_1 S(\theta_1, \theta_2)} \right\}$$

$$B(\theta_1, \theta_2) = - \left[\frac{g''(\theta_2)}{2g(\theta_2)} \right] + \frac{1}{2} \left[\frac{g'(\theta_2)}{g(\theta_2)} \right]^2 + n\theta_1 \left\{ \frac{\left[-2 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^3} + 3\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right]}{S(\theta_1, \theta_2)} \right\} -$$

$$- n\theta_1 \left\{ \frac{\left[\sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - \theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

onde $g'(\theta_2)$ e $g''(\theta_2)$ estão definidos em (E.2.1.4) e (E.2.1.5), respectivamente.

$$C(\theta_1, \theta_2) = \left\{ \frac{\left[-2n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} + n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} \right]}{S(\theta_1, \theta_2)} \right\} - \left\{ \frac{\left[n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

E.3.2. Média a posteriori para θ_2 :

O determinante da matriz Hessiana do numerador de (4.27) é:

$$\det(nD^2 h^*(\hat{\theta}_1^*, \hat{\theta}_2^*)) = F^*(\hat{\theta}_1, \hat{\theta}_2) G^*(\hat{\theta}_1, \hat{\theta}_2) - [H^*(\hat{\theta}_1, \hat{\theta}_2)]^2 \quad (\text{E.3.2.1})$$

onde:

$$F^*(\theta_1, \theta_2) = \frac{1}{\theta_1^2} + n \left\{ \frac{\sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2}}{S(\theta_1, \theta_2)} \right\} - \left\{ \frac{2 \left[\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right]}{\theta_1 S(\theta_1, \theta_2)} \right\}$$

$$G^*(\theta_1, \theta_2) = - \left[\frac{g''(\theta_2)}{2g(\theta_2)} \right] + \frac{1}{2} \left[\frac{g'(\theta_2)}{g(\theta_2)} \right]^2 + n\theta_1 \left\{ \frac{\left[-2 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^3} + 3\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right]}{S(\theta_1, \theta_2)} \right\} -$$

$$-n\theta_1 \left\{ \frac{\left[\sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - \theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

$$H^*(\theta_1, \theta_2) = \left\{ \frac{\left[-2n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} + n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} \right]}{S(\theta_1, \theta_2)} \right\} -$$

$$\left\{ \frac{\left[n\theta_1 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^2} - n \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)} \right] \cdot \left[2\theta_1 \sum_{i=1}^n \frac{x_i y_i}{(\theta_2 + x_i)^2} - 2\theta_1^2 \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]}{[S(\theta_1, \theta_2)]^2} \right\}$$

O determinante (E.3.2.1) da matriz Hessiana para o denominador de (4.26) coincide com o determinante,

$$\det(nD^2 h(\hat{\theta}_1, \hat{\theta}_2)) = F(\hat{\theta}_1, \hat{\theta}_2)G(\hat{\theta}_1, \hat{\theta}_2) - [H(\hat{\theta}_1, \hat{\theta}_2)]^2 \quad (\text{E.3.2.2})$$

da matriz Hessiana para o denominador de (4.27).

E.4. Expressões utilizadas na obtenção da densidade preditiva para y_{n+1}^* :

A expressão do determinante da matriz Hessiana de h , necessária para o cálculo da densidade preditiva para y_{n+1}^* , é dado por:

$$\det(nD^2 h(\hat{\theta}_1, \hat{\theta}_2)) = A(\hat{\theta}_1, \hat{\theta}_2)B(\hat{\theta}_1, \hat{\theta}_2) - [C(\hat{\theta}_1, \hat{\theta}_2)]^2 \quad (\text{E.4.1})$$

onde:

$$A(\theta_1, \theta_2) = \frac{1}{\theta_1^2} + (n+1) \left\{ \frac{\left[\sum_{i=1}^n \left(\frac{x_i^2}{(\theta_2 + x_i)^2} \right) + \frac{x_*^2}{(\theta_2 + x_*)^2} \right]}{\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x_*}{(\theta_2 + x_*)} \right)^2} \right\} -$$

$$-2(n+1) \left\{ \frac{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right) \cdot \left(\frac{-x_i}{(\theta_2 + x_i)} \right) + \left(y^* - \frac{\theta_1 x_*}{(\theta_2 + x_*)} \right) \cdot \left(\frac{-x_*}{(\theta_2 + x_*)} \right) \right]^2}{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x_*}{(\theta_2 + x_*)} \right)^2 \right]^2} \right\}$$

$$\begin{aligned}
B(\theta) = (n+1) & \left\{ \frac{\left[-2\theta_1 \sum_{i=1}^n \left(\frac{x_i y_i}{(\theta_2 + x_i)^3} \right) + 3\theta_1^2 \sum_{i=1}^n \left(\frac{x_i^2}{(\theta_2 + x_i)^4} \right) - 2\theta_1 \left(\frac{x^* y^*}{(\theta_2 + x^*)^3} \right) + 3\theta_1^2 \left(\frac{x^{*2}}{(\theta_2 + x^*)^4} \right) \right]}{\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x^*}{(\theta_1 + x^*)} \right)^2} \right\} - \\
& -2(n+1) \left\{ \frac{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right) \cdot \left(\frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) + \left(y^* - \frac{\theta_1 x^*}{(\theta_2 + x^*)} \right) \cdot \left(\frac{\theta_1 x^*}{(\theta_2 + x^*)^2} \right) \right]^2}{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x^*}{(\theta_1 + x^*)} \right)^2 \right]^2} \right\} \\
C(\theta_1, \theta_2) = (n+1) & \left\{ \frac{\left[\sum_{i=1}^n \left(\frac{x_i y_i}{(\theta_2 + x_i)^2} \right) - 2\theta_1 \sum_{i=1}^n \left(\frac{x_i^2}{(\theta_2 + x_i)^3} \right) + \left(\frac{x^* y^*}{(\theta_2 + x^*)^2} \right) - 2\theta_1 \left(\frac{x^{*2}}{(\theta_2 + x^*)^3} \right) \right]}{\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x^*}{(\theta_1 + x^*)} \right)^2} \right\} \\
& - (n+1) \left\{ \frac{2 \left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right) \cdot \left(\frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) + \left(y^* - \frac{\theta_1 x^*}{(\theta_2 + x^*)} \right) \cdot \left(\frac{\theta_1 x^*}{(\theta_2 + x^*)^2} \right) \right]}{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x^*}{(\theta_1 + x^*)} \right)^2 \right]^2} \right\} \times \\
& \times \left\{ \frac{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right) \cdot \left(\frac{-x_i}{(\theta_2 + x_i)} \right) + \left(y^* - \frac{\theta_1 x^*}{(\theta_2 + x^*)} \right) \cdot \left(\frac{-x^*}{(\theta_2 + x^*)} \right) \right]}{\left[\sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{(\theta_2 + x_i)} \right)^2 + \left(y^* - \frac{\theta_1 x^*}{(\theta_1 + x^*)} \right)^2 \right]^2} \right\}.
\end{aligned}$$

E.5. Expressões utilizadas na obtenção de densidade a posteriori marginais para β

E.5.1. Densidade a posteriori marginal para β_1 :

$$\begin{aligned}
 g'(\beta_1, \beta_2) &= 4 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2^3}{(\beta_1 + \beta_2 x_i)^5} \right) + \\
 &+ 2 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^4} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2}{(\beta_1 + \beta_2 x_i)^3} \right) - 6 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2^2}{(\beta_1 + \beta_2 x_i)^4} \right) \\
 g''(\beta_1, \beta_2) &= \left\{ 2 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^4} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1^2 - 2x_i^2 \beta_1 \beta_2}{(\beta_1 + \beta_2 x_i)^4} \right) + 8 \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2^3}{(\beta_1 + \beta_2 x_i)^5} \right) \right\} + \\
 &+ 4 \left\{ \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2^3}{(\beta_1 + \beta_2 x_i)^5} \right) \cdot \left(\sum_{i=1}^n \frac{2x_i^2 \beta_1 \beta_2}{(\beta_1 + \beta_2 x_i)^3} \right) + \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{(3x_i^2 \beta_1^2 \beta_2^2 - 2\beta_1 \beta_2^3 x_i^3)}{(\beta_1 + \beta_2 x_i)^6} \right) \right\} - \\
 &- 6 \left\{ \left(\sum_{i=1}^n \frac{x_i^2 \beta_1 \beta_2^2}{(\beta_1 + \beta_2 x_i)^4} \right) \cdot \left(\sum_{i=1}^n \frac{(3x_i^2 \beta_1 \beta_2^2)}{(\beta_1 + \beta_2 x_i)^4} \right) + \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{(2x_i^2 \beta_2 \beta_1^2 - 2\beta_1 \beta_2^2 x_i^3)}{(\beta_1 + \beta_2 x_i)^5} \right) \right\}
 \end{aligned}$$

E.5.2. Densidade a posteriori marginal para β_2 :

$$\begin{aligned}
 g'(\beta_1, \beta_2) &= -2 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^4} \right) - \\
 &- 4 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^5} \right) + 6 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^4} \right) \\
 g''(\beta_1, \beta_2) &= 6 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^4} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_1^4}{(\beta_1 + \beta_2 x_i)^4} \right) + 8 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^5} \right) + \\
 &+ 8 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^5} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^3} \right) + 20 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^2}{(\beta_1 + \beta_2 x_i)^2} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^4}{(\beta_1 + \beta_2 x_i)^6} \right) - \\
 &- 18 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^4} \right)^2 - 24 \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^3} \right) \cdot \left(\sum_{i=1}^n \frac{x_i^2 \beta_2^3}{(\beta_1 + \beta_2 x_i)^5} \right)
 \end{aligned}$$

Apêndice F

ALGUNS PROGRAMAS DESENVOLVIDOS

A seguir, apresentaremos dois programas desenvolvidos na linguagem *Fortran* e *Turbo Pascal*, respectivamente, utilizados para os cálculos em nossas análises.

1. Programa retirado de Ratkowsky (1983) para a obtenção das medidas de curvatura de Bates e Watts (1980), medida do vício de Box (1971) e estimativas dos parâmetros do modelo não linear via método de Gauss-Newton.

Inicialmente, apresentamos a subrotina externa e logo a seguir, o programa completo.

```

SUBROUTINE EVAL(PARAM,I,ITASK,WK1,IFAIL)
  INTEGER I,ITASK,IFAIL
  REAL PARAM(2),WK1(3),X,ZERO,ONE,TWO
  COMMON X(30)
  DATA ZERO/0.0E0/,ONE/1.0E0/,TWO/2.0E0/
  IFAIL=0
  IF(ITASK.NE.0) GO TO 1111
    WK1(1)=PARAM(1)*X(I)/(PARAM(2)+X(I))
    RETURN
1111 CONTINUE
  IF(ITASK.NE.1) GO TO 2222
    WK1(1)=X(I)/(PARAM(2)+X(I))
    WK1(2)=-PARAM(1)*X(I)/(PARAM(2)+X(I))**2
    RETURN
2222 CONTINUE
  IF(ITASK.NE.2) GO TO 3333
    WK1(1)=ZERO
    WK1(2)=-X(I)/(PARAM(2)+X(I))**2
    WK1(3)=2*X(I)*PARAM(1)/(PARAM(2)+X(I))**3
    RETURN
3333 CONTINUE
  IFAIL=1
  RETURN
END

EXTERNAL EVAL
  INTEGER JPVT(2),LUNIN,LUNOUT,NPTS,NP,NITS,ND1,I,J,JOB,NSIM,NS,
+   IFAIL
  REAL Y(30),PARAM(2),PAR(2),VARC(2,2),DERIVS(60),A(120),QRAUX(2),
+   WK1(3),WK2(30),DIFF(30),BIAS(2),X,ALPHA,
+   RSS,VAR,SD,FLOAT,SQRT,ALOG
  COMMON X(30)
  DATA PARAM(1)/5/,PARAM(2)/0.3/

```

```

DATA LUNIN/5/,LUNOUT/6/,NPTS/30/,NP/2/,NITS/30/,ND1/2/
DATA ALPHA/1.0E-6/
WRITE(LUNOUT,10)
DO 100 I=1,NPTS
    READ*, X(I), Y(I)
100 CONTINUE
DO I=1,NPTS
    PRINT*, X(I),Y(I)
END DO
JOB=11111
CALL SOLVE(Y,PARAM,ALPHA,NITS,NPTS,NP,ND1,LUNOUT,JOB,EVAL,
+   VARC,RSS,IFAIL,DERIVS,QRAUX,JPVT,DIFF,WK1)
IF(IFAIL.NE.0) STOP
JOB=111
VAR=RSS/FLOAT(NPTS-NP)
CALL BATES(PARAM,VAR,NITS,NPTS,NP,LUNOUT,JOB,EVAL,BIAS,IFAIL,
+   A,DERIVS,VARC,QRAUX,WK1,WK2,JPVT)
IF(IFAIL.NE.0) STOP
STOP
10  FORMAT(///1H ,10X,20HLEAF PRODUCTION DATA
+   //1H ,9X,1HX,9X,1HY//)
20  FORMAT(F5.1,F5.3)
30  FORMAT(1H ,5X,F5.1,5X,F5.3)
40  FORMAT(///1H ,10X,13HPARAMETER NO.,I4)
END

C
SUBROUTINE SOLVE(Y,PARAM,ALPHA,NITS,NPTS,NP,ND1,LUNOUT,JOB,EVAL,
+   VARC,RSS,IFAIL,DERIVS,QRAUX,JPVT,DIFF,WK1)
EXTERNAL EVAL
INTEGER JPVT(1),IFAIL,NPTS,NP,NITS,ND1,LUNOUT,JOB,I,J,K,L,ITER,
+   ITASK
REAL DERIVS(NPTS,1),VARC(ND1,1),Y(1),PARAM(1),DIFF(1),WK1(1),
+   QRAUX(1),DUM(1),ALPHA,RSS,RES,TEST,RESVAR,ACCUM,DENOM,ABS,
+   FLOAT,ZERO,ONE
LOGICAL CNVRGD,PRNPAR
DATA ZERO/0.0E0/,ONE/1.0E0/
IFAIL=0
PRNPAR=JOB/10000.NE.0
IF(.NOT.PRNPAR) GO TO 1111
    WRITE(LUNOUT,10) (I,I=1,NP)
    RSS=ZERO
    DO 100 I=1,NPTS
        ITASK=0
        CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
        IF(IFAIL.NE.0) GO TO 6666
        RES=Y(I)-WK1(1)
        RSS=RSS+RES*RES
100    CONTINUE
    WRITE(LUNOUT,20) RSS,(PARAM(I),I=1,NP)
1111 CONTINUE
ITER=1
2222 CONTINUE
    DO 200 I=1,NPTS
        ITASK=0
        CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
        IF(IFAIL.NE.0) GO TO 6666

```



```

        DIFF(I)=Y(I)-WK1(I)
        ITASK=1
        CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
        IF(IFAIL.NE.0) GO TO 6666
        DO 300 J=1,NP
            DERIVS(I,J)=WK1(J)
300      CONTINUE
200      CONTINUE
        DO 400 I=1,NP
            JPVT(I)=0
400      CONTINUE
        CALL SQDC(DERIVS,NPTS,NPTS,NP,QRAUX,JPVT,WK1,1)
        CALL SQSL(DERIVS,NPTS,NPTS,NP,QRAUX,DIFF,DUM,DIFF,WK1,DUM,
+          DUM,100,IFAIL)
        IF(IFAIL.NE.0) GO TO 5555
        CNVRGD=.TRUE.
        DO 500 I=1,NP
            J=JPVT(I)
            PARAM(J)=PARAM(J)+WK1(I)
            DENOM=PARAM(J)
            PRINT*, '*****', PARAM(J)
            IF(DENOM.EQ.ZERO) DENOM=ONE
            TEST=WK1(I)/DENOM
            IF(ABS(TEST).GT.ALPHA) CNVRGD=.FALSE.
500      CONTINUE
            RSS=ZERO
            DO 600 I=1,NPTS
                ITASK=0
                CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
                IF(IFAIL.NE.0) GO TO 6666
                RES=Y(I)-WK1(I)
                RSS=RSS+RES*RES
600      CONTINUE
            IF(.NOT.PRNP) GO TO 3333
            WRITE(LUNOUT,20) RSS,(PARAM(I),I=1,NP)
3333     CONTINUE
            ITER=ITER+1
            IF(ITER.GT.NITS) GO TO 4444
            IF(.NOT.CNVRGD) GO TO 2222
            IF(NPTS.EQ.NP) GO TO 7777
            CALL STRDI(DERIVS,NPTS,NP,DUM,011,IFAIL)
            IF(IFAIL.NE.0) GO TO 5555
            RESVAR=RSS/FLOAT(NPTS-NP)
            DO 700 I=1,NP
                DO 800 J=I,NP
                    ACCUM=ZERO
                    DO 900 K=J,NP
                        ACCUM=ACCUM+DERIVS(I,K)*DERIVS(J,K)
900          CONTINUE
                    ACCUM=ACCUM*RESVAR
                    K=JPVT(I)
                    L=JPVT(J)
                    VARC(K,L)=ACCUM
                    VARC(L,K)=ACCUM
800        CONTINUE
700      CONTINUE

```

```

      CALL DSPLAY(Y,VARC,PARAM,WK1,RESVAR,EVAL,NPTS,NP,ND1,LUNOUT,JOB,
+       IFAIL)
      IF(IFAIL.NE.0) GO TO 6666
      RETURN
4444  CONTINUE
      WRITE(LUNOUT,30) NITS
      IFAIL=1
      RETURN
5555  CONTINUE
      WRITE(LUNOUT,40)
      IFAIL=2
      RETURN
6666  CONTINUE
      WRITE(LUNOUT,50)
      IFAIL=8
      RETURN
7777  CONTINUE
      WRITE(LUNOUT,60)
      IFAIL=9
      RETURN
10    FORMAT(/1H ,13X,19HPARAMETER ESTIMATES/1H ,10X,3HRSS,9(I13)/(1H,
+       13X,9(I13)))
20    FORMAT(1H ,10(E13.6)/(1H ,13X,9(E13.6)))
30    FORMAT(/1H ,38H*** SUBROUTINE SOLVE CONVERGENCE NOT,
+       15H ACHIEVED AFTER,I10,2X,10HITERATIONS//)
40    FORMAT(/1H ,38H*** SUBROUTINE SOLVE SINGULAR MATRIX//)
50    FORMAT(/1H ,41H*** SUBROUTINE SOLVE USER-FLAGGED ERROR,
+       19H IN SUBROUTINE EVAL//)
60    FORMAT(/1H ,45H*** SUBROUTINE SOLVE RESIDUAL VARIANCE ZERO//)
      END
C
      SUBROUTINE BATES(PARAM,VAR,NITS,NPTS,NP,LUNOUT,JOB,EVAL,BIAS,
+       IFAIL,A,DERIVS,RINV,QRAUX,WK1,WK2,JPVT)
      EXTERNAL EVAL
      INTEGER JPVT(1),NITS,NPTS,NP,LUNOUT,JOB,IFAIL,I,J,K,NW,M,ITASK
      REAL PARAM(1),A(NPTS,NP,1),DERIVS(NPTS,1),RINV(NP,1),QRAUX(1),
+       BIAS(1),WK1(1),WK2(1),VAR,STRAD,TEMP,C1,C2,W,ZERO,SQRT,FLOAT
      DATA ZERO/0.0E0/
      IFAIL=0
      IF(VAR.LE.ZERO.OR.NP.LE.0) GO TO 2222
      STRAD=SQRT(VAR*FLOAT(NP))
      DO 100 I=1,NPTS
          ITASK=1
          CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
          IF(IFAIL.NE.0) GO TO 4444
          DO 200 J=1,NP
              DERIVS(I,J)=WK1(J)/STRAD
200      CONTINUE
          ITASK=2
          CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
          IF(IFAIL.NE.0) GO TO 4444
          NW=0
          DO 300 J=1,NP
              DO 400 K=1,J
                  NW=NW+1
                  W=WK1(NW)/STRAD

```

```

                                A(I,J,K)=W
                                IF(J.NE.K) A(I,K,J)=W
400      CONTINUE
300      CONTINUE
100      CONTINUE
        CALL ARRAYS(A,DERIVS,STRAD,NPTS,NP,LUNOUT,JOB)
        DO 700 I=1,NP
            JPVT(I)=0
700      CONTINUE
        CALL SQRDC(DERIVS,NPTS,NPTS,NP,QRAUX,JPVT,WK1,1)
        DO 500 I=1,NP
            IF(QRAUX(I).EQ.ZERO.OR.DERIVS(I,I).EQ.ZERO) GO TO 3333
            TEMP=QRAUX(I)
            QRAUX(I)=DERIVS(I,I)
            DERIVS(I,I)=TEMP
            DO 600 J=I,NPTS
                DERIVS(J,I)=-DERIVS(J,I)*QRAUX(I)
600      CONTINUE
500      CONTINUE
        CALL ACCEL(A,DERIVS,RINV,QRAUX,WK1,JPVT,NPTS,NP)
        CALL BOX(A,RINV,BIAS,WK1,NPTS,NP)
        M=NP+1
        CALL CURVE(A,DERIVS,WK1,QRAUX,WK2,C1,M,NPTS,NITS,NPTS,NP,IFAIL)
        IF(IFAIL.EQ.0) GO TO 1111
            IFAIL=5
            WRITE(LUNOUT,30) NITS
1111     CONTINUE
        CALL CURVE(A,DERIVS,WK1,QRAUX,WK2,C2,1,NP,NITS,NPTS,NP,IFAIL)
        IF(IFAIL.EQ.6) WRITE(LUNOUT,30) NITS
        CALL WATTS(A,BIAS,PARAM,C1,C2,NPTS,NP,LUNOUT,JOB)
        RETURN
2222     CONTINUE
        IFAIL=3
        WRITE(LUNOUT,10)
        RETURN
3333     CONTINUE
        IFAIL=4
        WRITE(LUNOUT,20)
        RETURN
4444     CONTINUE
        IFAIL=8
        WRITE(LUNOUT,40)
        RETURN
10      FORMAT(/1H ,48H*** SUBROUTINE BATES INCORRECT STANDARD RADIUS
+           //)
20      FORMAT(/1H ,38H*** SUBROUTINE BATES  MATRIZ SINGULAR //)
30      FORMAT(/1H ,38H*** SUBROUTINE BATES  CONVERGENCE NOT,
+           15H ACHIEVED AFTER,I10,2X,10HITERATIONS//)
40      FORMAT(/1H ,41H*** SUBROUTINE BATES  USER-FLAGGED ERROR,
+           19H IN SUBROUTINE EVAL//)
        END
C
        SUBROUTINE ACCEL(A,DERIVS,RINV,ALPHA,TEMPV,JPVT,NPTS,NP)
        INTEGER JPVT(1),NPTS,NP,I,J,K,L,IMIN1,JJ,JPLUS1
        REAL A(NPTS,NP,1),DERIVS(NPTS,1),RINV(NP,1),ALPHA(1),
+           TEMPV(1),ACCUM,GAMMA,ZERO,ONE

```



```

DATA ZERO/0.0E0/,ONE/1.0E0/
DO 100 I=1,NP
  DO 200 J=1,I
    DO 300 K=1,NP
      ACCUM=ZERO
      DO 400 L=K,NPTS
        ACCUM=ACCUM+DERIVS(L,K)*A(L,I,J)
400      CONTINUE
      GAMMA=ACCUM/(ALPHA(K)*DERIVS(K,K))
      DO 500 L=K,NPTS
        A(L,I,J)=A(L,I,J)+GAMMA*DERIVS(L,K)
500      CONTINUE
300    CONTINUE
    DO 600 K=1,NPTS
      A(K,J,I)=A(K,I,J)
600    CONTINUE
200  CONTINUE
100 CONTINUE
DO 700 I=1,NP
  DO 800 J=1,NP
    TEMPV(J)=ZERO
800  CONTINUE
    TEMPV(I)=ONE/ALPHA(I)
    IMIN1=I-1
    IF(IMIN1.LT.1) GO TO 1111
    DO 900 JJ=1,IMIN1
      J=IMIN1-JJ+1
      JPLUS1=J+1
      ACCUM=ZERO
      DO 1000 K=JPLUS1,NP
        ACCUM=ACCUM+DERIVS(J,K)*TEMPV(K)
1000     CONTINUE
        TEMPV(J)=-ACCUM/ALPHA(J)
900     CONTINUE
1111 CONTINUE
    DO 1100 J=1,NP
      K=JPVT(J)
      RINV(K,I)=TEMPV(J)
1100 CONTINUE
700 CONTINUE
DO 1200 I=1,NPTS
  DO 1300 J=1,NP
    DO 1400 K=1,NP
      ACCUM=ZERO
      DO 1500 L=1,NP
        ACCUM=ACCUM+A(I,J,L)*RINV(L,K)
1500     CONTINUE
        DERIVS(J,K)=ACCUM
1400     CONTINUE
1300 CONTINUE
    DO 1600 J=1,NP
      DO 1700 K=1,J
        ACCUM=ZERO
        DO 1800 L=1,NP
          ACCUM=ACCUM+RINV(L,J)*DERIVS(L,K)
1800     CONTINUE

```

```

                A(I,J,K)=ACCUM
                A(I,K,J)=ACCUM
1700      CONTINUE
1600      CONTINUE
1200      CONTINUE
          RETURN
          END

C
  SUBROUTINE ARRAYS(A,DERIVS,STRAD,NPTS,NP,LUNOUT,JOB)
    INTEGER NPTS,NP,JOB,LUNOUT,I,J,K,MOD
    REAL A(NPTS,NP,1),DERIVS(NPTS,1),STRAD
    LOGICAL PRND1,PRND2
    PRND1=JOB/10000.NE.0
    PRND2=MOD(JOB,10000)/1000.NE.0
    IF((.NOT.PRND1).AND.(.NOT.PRND2)) GO TO 1111
    WRITE(LUNOUT,10) STRAD
1111      CONTINUE
          IF(.NOT.PRND1) GO TO 2222
                WRITE(LUNOUT,20)
                DO 100 I=1,NPTS
                        WRITE(LUNOUT,30) I,(DERIVS(I,J),J=1,NP)
100      CONTINUE
2222      CONTINUE
          IF(.NOT.PRND2) GO TO 3333
                WRITE(LUNOUT,40)
                DO 200 I=1,NP
                        WRITE(LUNOUT,50) I
                        DO 300 J=1,NP
                                WRITE(LUNOUT,60) (A(I,J,K),K=1,NP)
300      CONTINUE
200      CONTINUE
3333      CONTINUE
          RETURN
10      FORMAT(///1H ,10X,17HSTANDARD RADIUS =,E13.6)
20      FORMAT(///1H ,10X,31HFIRST-ORDER PARTIAL DERIVATIVES,
+            15H (STANDARDIZED)//1H ,10HDATA POINT)
30      FORMAT(1H ,110,10X,8E12.5/(1H ,20X,8E12.5))
40      FORMAT(///1H ,10X,32HSECOND-ORDER PARTIAL DERIVATIVES,
+            15H (STANDARDIZED)//1H ,10HDATA POINT)
50      FORMAT(1H ,15)
60      FORMAT(1H ,10X,8E12.5/(1H ,10X,8E12.5))
          END

C
  SUBROUTINE BOX(A,RINV,BIAS,TEMPV,NPTS,NP)
    INTEGER NPTS,NP,I,J
    REAL A(NPTS,NP,1),RINV(NP,1),BIAS(1),TEMPV(1),ZERO,Float
    DATA ZERO/0.0E0/
    DO 100 I=1,NP
            TEMPV(I)=ZERO
            DO 200 J=1,NP
                    TEMPV(I)=TEMPV(I)+A(I,J,J)
200      CONTINUE
            TEMPV(I)=-TEMPV(I)/Float(2*NP)
100      CONTINUE
          DO 300 I=1,NP
                BIAS(I)=ZERO

```

```

        DO 400 J=1,NP
            BIAS(I)=BIAS(I)+RINV(I,J)*TEMPV(J)
400      CONTINUE
300    CONTINUE
        RETURN
        END

C
    SUBROUTINE CNTEST(DIR,GRAD,COSA,SINA,THQA,NP,CNVGED)
    INTEGER NP,I
    REAL DIR(1),GRAD(1),COSA,SINA,THQA,EPS,PI,TEST,ABS,SIGN,ATAN2,
+      SGN,ZERO,ONE,TWO,PSF,SQRT
    LOGICAL CNVGED
    DATA EPS/1.0E-4/,PI/3.14159265E0/,ZERO/0.0E0/,ONE/1.0E0/,
+      TWO/2.0E0/,PSF/7.5E-1/
    CNVGED=.FALSE.
    TEST=ONE-EPS
    IF(ABS(COSA).LE.TEST) GO TO 1111
        SGN=SIGN(ONE,COSA)
        DO 100 I=1,NP
            DIR(I)=SGN*GRAD(I)
100      CONTINUE
        CNVGED=.TRUE.
        RETURN
1111   CONTINUE
        SINA=SQRT(ONE-COSA**2)
        IF(SINA.NE.ONE) GO TO 2222
        THQA=PSF*PI/TWO
        RETURN
2222   CONTINUE
        IF(COSA.GE.ZERO) GO TO 3333
        THQA=PSF*(ATAN2(SINA,COSA)+PI)
        RETURN
3333   CONTINUE
        THQA=PSF*(ATAN2(SINA,COSA))
        RETURN
        END

C
    SUBROUTINE CURVE(A,TEMPML,DIR,GRAD,TEMPL,CMAX,IM,IN,NITS,NPTS,
+      NP,IFAIL)
    INTEGER IM,IN,NITS,NPTS,NP,IFAIL,I,J,K,ITER,NPMIN1
    REAL A(NPTS,NP,1),TEMPML(NPTS,1),DIR(1),GRAD(1),TEMPL(1),
+      CMAX,ACCUM,ALENG,SINA,COSA,SIN,COS,SQRT,THQA,C1,C2,ZERO,ONE
    LOGICAL CNVGED
    DATA ZERO/0.0E0/,ONE/1.0E0/
    NPMIN1=NP-1
    IF(NPMIN1.LE.1) GO TO 1111
        DO 400 I=1,NPMIN1
            DIR(I)=ZERO
400      CONTINUE
1111   CONTINUE
        DIR(NP)=ONE
        ITER=1
2222   CONTINUE
        DO 500 I=IM,IN
            DO 600 J=1,NP
                ACCUM=ZERO

```



```

                DO 700 K=1,NP
                    ACCUM=ACCUM+A(I,J,K)*DIR(K)
700             CONTINUE
                TEMPML(I,J)=ACCUM
600             CONTINUE
500             CONTINUE
                DO 800 I=IM,IN
                    ACCUM=ZERO
                    DO 900 J=1,NP
                        ACCUM=ACCUM+TEMPML(I,J)*DIR(J)
900             CONTINUE
                TEMPL(I)=ACCUM
800             CONTINUE
                DO 1000 I=1,NP
                    ACCUM=ZERO
                    DO 1100 J=IM,IN
                        ACCUM=ACCUM+TEMPML(J,I)*TEMPL(J)
1100            CONTINUE
                GRAD(I)=ACCUM
1000            CONTINUE
                ACCUM=ZERO
                DO 1200 I=1,NP
                    ACCUM=ACCUM+GRAD(I)*GRAD(I)
1200            CONTINUE
                ALENG=SQRT(ACCUM)
                IF(ALENG.EQ.ZERO) GO TO 4444
                DO 1300 I=1,NP
                    GRAD(I)=GRAD(I)/ALENG
1300            CONTINUE
                COSA=ZERO
                DO 1400 I=1,NP
                    COSA=COSA+DIR(I)*GRAD(I)
1400            CONTINUE
                CALL CNTEST(DIR,GRAD,COSA,SINA,THQA,NP,CNVGED)
                IF(CNVGED) GO TO 4444
                    ITER=ITER+1
                    IF(ITER.LE.NITS) GO TO 3333
                        IFAIL=6
                        GO TO 4444
3333            CONTINUE
                C2=SIN(THQA)/SINA
                C1=COS(THQA)-C2*COSA
                DO 1500 I=1,NP
                    DIR(I)=C1*DIR(I)+C2*GRAD(I)
1500            CONTINUE
                GO TO 2222
4444            CONTINUE
                DO 1600 I=IM,IN
                    DO 1700 J=1,NP
                        ACCUM=ZERO
                        DO 1800 K=1,NP
                            ACCUM=ACCUM+A(I,J,K)*DIR(K)
1800            CONTINUE
                        TEMPML(I,J)=ACCUM
1700            CONTINUE
1600            CONTINUE

```

```

DO 1900 I=IM,IN
  ACCUM=ZERO
  DO 2000 J=1,NP
    ACCUM=ACCUM+TEMPML(I,J)*DIR(J)
2000  CONTINUE
    TEMPL(I)=ACCUM
1900  CONTINUE
    ACCUM=ZERO
    DO 2100 I=IM,IN
      ACCUM=ACCUM+TEMPL(I)*TEMPL(I)
2100  CONTINUE
    CMAX=SQRT(ACCUM)
    RETURN
    END
C
  SUBROUTINE DSPLAY(Y,VARC,PARAM,WK1,VAR,EVAL,NPTS,NP,ND1,
+    LUNOUT,JOB,IFAIL)
  EXTERNAL EVAL
  INTEGER JOB,MOD,LUNOUT,ND1,NP,I,J,NPTS,IFAIL,ITASK
  REAL Y(1),VARC(ND1,1),PARAM(1),WK1(1),VAR,SQRT,T,SE,YHAT,RES
  LOGICAL PRNTAB, PRNVCV,PRNCOR,PRNFIT
  PRNTAB=MOD(JOB,10000)/1000.NE.0
  PRNVCV=MOD(JOB,1000)/100.NE.0
  PRNCOR=MOD(JOB,100)/10.NE.0
  PRNFIT=MOD(JOB,10).NE.0
  IF(.NOT.PRNTAB) GO TO 1111
  WRITE(LUNOUT,10)
  DO 100 I=1,NP
    SE=SQRT(VARC(I,I))
    T=PARAM(I)/SE
    WRITE(LUNOUT,20) I,PARAM(I),SE,T
100  CONTINUE
1111  CONTINUE
  IF(.NOT.PRNVCV) GO TO 2222
  WRITE(LUNOUT,30)
  DO 200 I=1,NP
    WRITE(LUNOUT,40) I,(VARC(I,J),J=1,I)
200  CONTINUE
2222  CONTINUE
  IF(.NOT.PRNCOR) GO TO 3333
  WRITE(LUNOUT,50)
  DO 300 I=1,NP
    DO 400 J=1,I
      WK1(J)=VARC(I,J)/SQRT(VARC(I,I)*VARC(J,J))
400  CONTINUE
    WRITE(LUNOUT,60) I,(WK1(J),J=1,I)
300  CONTINUE
3333  CONTINUE
  IF(.NOT.PRNFIT) GO TO 4444
  WRITE(LUNOUT,70)
  DO 500 I=1,NPTS
    ITASK=0
    CALL EVAL(PARAM,I,ITASK,WK1,IFAIL)
    IF(IFAIL.NE.0) RETURN
    YHAT=WK1(1)
    RES=Y(I)-YHAT

```

```

WRITE(LUNOUT,80) I,Y(I),YHAT,RES
500  CONTINUE
4444 CONTINUE
      RETURN
10   FORMAT(////1H,10X,34HPARAMETER ESTIMATES AT CONVERGENCE/
+       /1H,9HPARAMETER,4X,8HESTIMATE,11X,2HSE,11X,1HT)
20   FORMAT(1H,15,4X,E12.6,1X,E12.6,2X,F10.2)
30   FORMAT(////1H,10X,36HPARAMETER VARIANCE-COVARIANCE MATRIX//)
40   FORMAT(1H,15,8(2X,E12.6)/(1H,5X,8(2X,E12.6)))
50   FORMAT(////1H,10X,22HPARAMETER CORRELATIONS//)
60   FORMAT(1H,15,8(2X,F12.6)/(1H,5X,8(2X,F12.6)))
70   FORMAT(////1H,11X,4HUNIT,14X,1HY,9X,6HFITTED,7X,8HRESIDUAL)
80   FORMAT(1H,115,3E15.6)
      END

C
      SUBROUTINE SORT(A,II,JJ)
      INTEGER IU(16),IL(16),I,II,J,JJ,IJ,K,M,L
      REAL A(1),T,TT
      M=1
      I=II
      J=JJ
5     IF(I.GE.J) GO TO 70
10    K=I
      IJ=(J+I)/2
      T=A(IJ)
      IF(A(I).LE.T) GO TO 20
      A(IJ)=A(I)
      A(I)=T
      T=A(IJ)
20    L=J
      IF(A(J).GE.T) GO TO 40
      A(IJ)=A(J)
      A(J)=T
      T=A(IJ)
      IF(A(I).LE.T) GO TO 40
      A(IJ)=A(I)
      A(I)=T
      T=A(IJ)
      GO TO 40
30    A(L)=A(K)
      A(K)=TT
40    L=L-1
      IF(A(L).GT.T) GO TO 40
      TT=A(L)
50    K=K+1
      IF(A(K).LT.T) GO TO 50
      IF(K.LE.L) GO TO 30
      IF(L-I.LE.J-K) GO TO 60
      IL(M)=I
      IU(M)=L
      I=K
      M=M+1
      GO TO 80
60    IL(M)=K
      IU(M)=J
      J=L

```



```

      M=M+1
      GO TO 80
70    M=M-1
      IF(M.EQ.0) RETURN
      I=IL(M)
      J=IU(M)
80    IF(J-I.GE.11) GO TO 10
      IF(I.EQ.II) GO TO 5
      I=I-1
90    I=I+1
      IF(I.EQ.J) GO TO 70
      T=A(I+1)
      IF(A(I).LE.T) GO TO 90
      K=I
100   A(K+1)=A(K)
      K=K-1
      IF(T.LT.A(K)) GO TO 100
      A(K+1)=T
      GO TO 90
      END

```

C

```

      SUBROUTINE WATTS(A,BIAS,PARAM,C1,C2,NPTS,NP,LUNOUT,JOB)
      INTEGER NPTS,NP,LUNOUT,JOB,MOD,I,J,K
      REAL A(NPTS,NP,1),BIAS(1),PARAM(1),C1,C2,PERCEN,ZERO,HUNDRD
      LOGICAL PRNC,PRNB,PRNA
      DATA ZERO/0.0E0/,HUNDRD/1.0E2/
      PRNC=MOD(JOB,10).NE.0
      PRNB=MOD(JOB,100)/10.NE.0
      PRNA=MOD(JOB,1000)/100.NE.0
      IF(.NOT.PRNA) GO TO 1111
      WRITE(LUNOUT,10)
      DO 200 I=1,NP
         WRITE(LUNOUT,20)
         DO 300 J=1,NP
            WRITE(LUNOUT,30) (A(I,J,K),K=1,NP)
300        CONTINUE
200       CONTINUE
1111      CONTINUE
      IF(.NOT.PRNB) GO TO 3333
      WRITE(LUNOUT,40)
      DO 400 I=1,NP
         IF(PARAM(I).EQ.ZERO) GO TO 2222
         PERCEN=BIAS(I)/PARAM(I)*HUNDRD
         WRITE(LUNOUT,50) I,PARAM(I),BIAS(I),PERCEN
         GO TO 400
2222      CONTINUE
         WRITE(LUNOUT,50) I,PARAM(I),BIAS(I)
400       CONTINUE
3333      CONTINUE
      IF(.NOT.PRNC) GO TO 4444
      WRITE(LUNOUT,60) C1,C2
4444      CONTINUE
      RETURN
10    FORMAT(//1H ,10X,18HACCELERATION ARRAY,
+          8H (PARAMETER-EFFECTS PORTION))
20    FORMAT(/1H )

```

```

30  FORMAT(1H,10X,10F12.4)
40  FORMAT(/1H,10X,10HBOX'S BIAS///1H,8X,9HPARAMETER,9X,
+    11HLS ESTIMATE,16X,4HBIAS,8X,12HPERCENT BIAS/)
50  FORMAT(1H,6X,9HPARAMETER,I2,2E20.8,F20.4)
60  FORMAT(/1H,10X,26HMAXIMUM CURVATURE MEASURES//1H,
+    14HINTRINSIC (IN),11X,F12.4//1H,22HPARAMETER-EFFECTS (PE),
+    3X,F12.4)
    END

```

C

```

SUBROUTINE SQRDC(X,LDX,N,P,QRAUX,JPVT,WORK,JOB)
INTEGER LDX,N,P,JOB
INTEGER JPVT(1)
REAL X(LDX,1),QRAUX(1),WORK(1)
INTEGER J,JP,L,LP1,LUP,MAXJ,PL,PU
REAL MAXNRM,SNRM2,TT
REAL SDOT,NRMXL,T
LOGICAL NEGJ,SWAPJ
PL=1
PU=0
IF (JOB.EQ.0) GO TO 60
DO 20 J=1,P
  SWAPJ=JPVT(J).GT.0
  NEGJ=JPVT(J).LT.0
  JPVT(J)=J
  IF(NEGJ) JPVT(J) = -J
  IF(.NOT.SWAPJ) GO TO 10
  IF(J.NE.PL) CALL SSWAP(N,X(1,PL),1,X(1,J),1)
  JPVT(J)=JPVT(PL)
  JPVT(PL)=J
  PL=PL+1
10  CONTINUE
20  CONTINUE
  PU=P
  DO 50 JJ=1,P
    J=P-JJ+1
    IF(JPVT(J).GE.0) GO TO 40
    JPVT(J)=-JPVT(J)
    IF (J.EQ.PU) GO TO 30
    CALL SSWAP(N,X(1,PU),1,X(1,J),1)
    JP=JPVT(PU)
    JPVT(PU)=JPVT(J)
    JPVT(J)=JP
30  CONTINUE
    PU=PU-1
40  CONTINUE
50  CONTINUE
60  CONTINUE
    IF(PU.LT.PL) GO TO 80
    DO 70 J=PL,PU
      QRAUX(J)=SNRM2(N,X(1,J),1)
      WORK(J)=QRAUX(J)
70  CONTINUE
80  CONTINUE
    LUP=MIN0(N,P)
    DO 200 L=1,LUP
      IF(L.LT.PL.OR.L.GE.PU) GO TO 120

```

```

      MAXNRM=0.0E0
      MAXJ=L
      DO 100 J=L,PU
      IF(QRAUX(J).LE.MAXNRM) GO TO 90
      MAXNRM=QRAUX(J)
      MAXJ=J
90    CONTINUE
100   CONTINUE
      IF(MAXJ.EQ.L) GO TO 110
      CALL SSWAP(N,X(1,L),1,X(1,MAXJ),1)
      QRAUX(MAXJ)=QRAUX(L)
      WORK(MAXJ)=WORK(L)
      JP=JPVT(MAXJ)
      JPVT(MAXJ)=JPVT(L)
      JPVT(L)=JP
110   CONTINUE
120   CONTINUE
      QRAUX(L)=0.0E0
      IF(L.EQ.N) GO TO 190
      NRMXL=SNRM2(N-L+1,X(L,L),1)
      IF (NRMXL.EQ.0.0E0) GO TO 180
      IF (X(L,L).NE.0.0E0) NRMXL=SIGN(NRMXL,X(L,L))
      CALL SSCAL(N-L+1,1.0E0/NRMXL,X(L,L),1)
      X(L,L)=1.0E0+X(L,L)
      LP1=L+1
      IF(P.LT.LP1) GO TO 170
      DO 160 J=LP1,P
      T=-SDOT(N-L+1,X(L,L),1,X(L,J),1)/X(L,L)
      CALL SAXPY(N-L+1,T,X(L,L),1,X(L,J),1)
      IF(J.LT.PL.OR.J.GT.PU) GO TO 150
      IF(QRAUX(J).EQ.0.0E0) GO TO 150
      TT=1.0E0-(ABS(X(L,J))/QRAUX(J))**2
      TT=AMAX1(TT,0.0E0)
      T=TT
      TT=1.0E0+0.05E0*TT*(QRAUX(J)/WORK(J))**2
      IF(TT.EQ.1.0E0) GO TO 130
      QRAUX(J)=QRAUX(J)*SQRT(T)
      GO TO 140
130   CONTINUE
      QRAUX(J)=SNRM2(N-L,X(L+1,J),1)
      WORK(J)=QRAUX(J)
140   CONTINUE
150   CONTINUE
160   CONTINUE
170   CONTINUE
      QRAUX(L)=X(L,L)
      X(L,L)=-NRMXL
180   CONTINUE
190   CONTINUE
200   CONTINUE
      RETURN
      END
C
      SUBROUTINE SQRSL(X,LDX,N,K,QRAUX,Y,QY,QTY,B,RSD,XB,JOB,INFO)
      INTEGER LDX,N,K,JOB,INFO
      REAL X(LDX,1),QRAUX(1),Y(1),QY(1),QTY(1),B(1),RSD(1),XB(1)

```



```

INTEGER I,J,II,JU,KP1
REAL SDOT,T,TEMP
LOGICAL CB,CQY,CQTY,CR,CXB
INFO=0
CQY=JOB/10000.NE.0
CQTY=MOD(JOB,10000).NE.0
CB=MOD(JOB,1000)/100.NE.0
CR=MOD(JOB,100)/10.NE.0
CXB=MOD(JOB,10).NE.0
JU=MIN0(K,N-1)
IF(JU.NE.0) GO TO 40
IF(CQY) QY(1)=Y(1)
IF(CQTY) QTY(1)=Y(1)
IF(CXB) XB(1)=Y(1)
IF(.NOT.CB) GO TO 30
IF(X(1,1).NE.0.0E0) GO TO 10
INFO=1
GO TO 20
10  CONTINUE
    B(1)=Y(1)/X(1,1)
20  CONTINUE
30  CONTINUE
    IF(CR) RSD(1)=0.0E0
    GO TO 250
40  CONTINUE
    IF(CQY) CALL SCOPY(N,Y,1,QY,1)
    IF(CQTY) CALL SCOPY(N,Y,1,QTY,1)
    IF(.NOT.CQY) GO TO 70
    DO 60 JJ=1,JU
        J=JU-JJ+1
        IF(QRAUX(J).EQ.0.0E0) GO TO 50
        TEMP=X(J,J)
        X(J,J)=QRAUX(J)
        T=-SDOT(N-J+1,X(J,J),1,QY(J),1)/X(J,J)
        CALL SAXPY(N-J+1,T,X(J,J),1,QY(J),1)
        X(J,J)=TEMP
50  CONTINUE
60  CONTINUE
70  CONTINUE
    IF(.NOT.CQTY) GO TO 100
    DO 90 J=1,JU
        IF(QRAUX(J).EQ.0.0E0) GO TO 80
        TEMP=X(J,J)
        X(J,J)=QRAUX(J)
        T=-SDOT(N-J+1,X(J,J),1,QTY(J),1)/X(J,J)
        CALL SAXPY(N-J+1,T,X(J,J),1,QTY(J),1)
        X(J,J)=TEMP
80  CONTINUE
90  CONTINUE
100 CONTINUE
    IF(CB) CALL SCOPY(K,QTY,1,B,1)
    KP1=K+1
    IF(CXB) CALL SCOPY(K,QTY,1,XB,1)
    IF(CR.AND.K.LT.N) CALL SCOPY(N-K,QTY(KP1),1,RSD(KP1),1)
    IF(.NOT.CXB.OR.KP1.GT.N) GO TO 120
    DO 110 I=KP1,N

```

```

XB(I)=0.0E0
110 CONTINUE
120 CONTINUE
    IF(.NOT.CR) GO TO 140
    DO 130 I=1,K
    RSD(I)=0.0E0
130 CONTINUE
140 CONTINUE
    IF(.NOT.CB) GO TO 190
    DO 170 JJ=1,K
    J=K-JJ+1
    IF(X(J,J).NE.0.0E0) GO TO 150
    INFO=J
    GO TO 180
150 CONTINUE
    B(J)=B(J)/X(J,J)
    IF(J.EQ.1) GO TO 160
    T=-B(J)
    CALL SAXPY(J-1,T,X(1,J),1,B,1)
160 CONTINUE
170 CONTINUE
180 CONTINUE
190 CONTINUE
    IF(.NOT.CR .AND. .NOT.CXB) GO TO 240
    DO 230 JJ=1,JU
    J=JU-JJ+1
    IF(QRAUX(J).EQ.0.0E0) GO TO 220
    TEMP=X(J,J)
    X(J,J)=QRAUX(J)
    IF(.NOT.CR) GO TO 200
    T=-SDOT(N-J+1,X(J,J),1,RSD(J),1)/X(J,J)
    CALL SAXPY(N-J+1,T,X(J,J),1,RSD(J),1)
200 CONTINUE
    IF(.NOT.CXB) GO TO 210
    T=-SDOT(N-J+1,X(J,J),1,XB(J),1)/X(J,J)
    CALL SAXPY(N-J+1,T,X(J,J),1,XB(J),1)
210 CONTINUE
    X(J,J)=TEMP
220 CONTINUE
230 CONTINUE
240 CONTINUE
250 CONTINUE
    RETURN
    END

```

C

```

SUBROUTINE SSWAP (N,SX,INCX,SY,INCY)
REAL SX(1),SY(1),STEMP
INTEGER I,INCX,INCY,IX,IY,M,MPI,N
IF(N.LE.0) RETURN
IF(INCX.EQ.1.AND.INCY.EQ.1) GO TO 20
IX = 1
IY = 1
IF(INCX.LT.0)IX = (-N+1)*INCX + 1
IF(INCY.LT.0)IY = (-N+1)*INCY + 1
DO 10 I = 1,N
STEMP = SX(IX)

```

```

      SX(IX) = SY(IY)
      SY(IY) = STEMP
      IX = IX + INCX
      IY = IY + INCY
10    CONTINUE
      RETURN
20    M = MOD(N,3)
      IF(M.EQ.0) GO TO 40
      DO 30 I = 1,M
        STEMP = SX(I)
        SX(I) = SY(I)
        SY(I) = STEMP
30    CONTINUE
      IF(N.LT.3) RETURN
40    MP1 = M + 1
      DO 50 I = MP1,N,3
        STEMP = SX(I)
        SX(I) = SY(I)
        SY(I) = STEMP
        STEMP = SX(I+1)
        SX(I+1) = SY(I+1)
        SY(I+1) = STEMP
        STEMP = SX(I+2)
        SX(I+2) = SY(I+2)
        SY(I+2) = STEMP
50    CONTINUE
      RETURN
      END

C
      SUBROUTINE SSCAL(N,SA,SX,INCX)
      REAL SA,SX(1)
      INTEGER I,INCX,M,MP1,N,NINCX
      IF(N.LE.0) RETURN
      IF(INCX.EQ.1) GO TO 20
      NINCX = N*INCX
      DO 10 I = 1,NINCX,INCX
        SX(I) = SA*SX(I)
10    CONTINUE
      RETURN
20    M = MOD(N,5)
      IF(M.EQ.0) GO TO 40
      DO 30 I = 1,M
        SX(I) = SA*SX(I)
30    CONTINUE
      IF(N.LT.5) RETURN
40    MP1 = M+1
      DO 50 I=MP1,N,5
        SX(I) = SA*SX(I)
        SX(I+1) = SA*SX(I+1)
        SX(I+2) = SA*SX(I+2)
        SX(I+3) = SA*SX(I+3)
        SX(I+4) = SA*SX(I+4)
50    CONTINUE
      RETURN
      END

C

```



```

REAL FUNCTION SNRM2(N,SX,INCX)
  INTEGER NEXT
  REAL SX(1),CUTLO,CUTHI,HITEST,SUM,XMAX,ZERO,ONE
  DATA ZERO,ONE/0.0E0, 1.0E0/
  DATA CUTLO, CUTHI / 4.441E-16, 1.304E19 /
  IF(N.GT.0) GO TO 10
  SNRM2 = ZERO
  GO TO 300
10  ASSIGN 30 TO NEXT
  SUM = ZERO
  NN = N*INCX
  I=1
20  GO TO NEXT,(30,50,70,110)
30  IF(ABS(SX(I)).GT.CUTLO) GO TO 85
  ASSIGN 50 TO NEXT
  XMAX = ZERO
50  IF(SX(I).EQ.ZERO) GO TO 200
  IF(ABS(SX(I)).GT.CUTLO) GO TO 85
  ASSIGN 70 TO NEXT
  GO TO 105
100 I=J
  ASSIGN 110 TO NEXT
  SUM=(SUM / SX(I)) / SX(I)
105 XMAX = ABS(SX(I))
  GO TO 115
70  IF(ABS(SX(I)).GT.CUTLO) GO TO 75
110 IF(ABS(SX(I)).LE.XMAX) GO TO 115
  SUM = ONE + SUM*(XMAX / SX(I))**2
  XMAX = ABS(SX(I))
  GO TO 200
115 SUM = SUM + (SX(I)/XMAX)**2
  GO TO 200
75  SUM = (SUM*XMAX)*XMAX
85  HITEST = CUTHI/FLOAT(N)
  DO 95 J=I,NN,INCX
  IF(ABS(SX(J)).GE.HITEST) GO TO 100
95  SUM = SUM+SX(J)**2
  SNRM2 = SQRT(SUM)
  GO TO 300
200 CONTINUE
  I=I+INCX
  IF(I.LE.NN) GO TO 20
  SNRM2 = XMAX*SQRT(SUM)
300 CONTINUE
  RETURN
  END

```

C

```

REAL FUNCTION SDOT(N,SX,INCX,SY,INCY)
  REAL SX(1),SY(1),STEMP
  INTEGER I,INCX,INCY,IX,IY,M,MP1,N
  STEMP = 0.0E0
  SDOT = 0.0E0
  IF(N.LE.0) RETURN
  IF(INCX.EQ.1.AND.INCY.EQ.1) GO TO 20
  IX = 1
  IY = 1

```

```

      IF(INCX.LT.0)IX = (-N+1)*INCX + 1
      IF(INCY.LT.0)IY = (-N+1)*INCY + 1
      DO 10 I=1,N
      STEMP = STEMP+SX(IX)*SY(IY)
      IX = IX + INCX
      IY = IY + INCY
10    CONTINUE
      SDOT = STEMP
      RETURN
20    M = MOD(N,5)
      IF(M.EQ.0) GO TO 40
      DO 30 I = 1,M
      STEMP = STEMP + SX(I)*SY(I)
30    CONTINUE
      IF(N.LT.5) GO TO 60
40    MP1 = M + 1
      DO 50 I = MP1,N,5
      STEMP = STEMP+SX(I)*SY(I)+SX(I+1)*SY(I+1)+SX(I+2)*SY(I+2)+
+      SX(I+3)*SY(I+3)+SX(I+4)*SY(I+4)
50    CONTINUE
60    SDOT = STEMP
      RETURN
      END

```

C

```

      SUBROUTINE SCOPY(N,SX,INCX,SY,INCY)
      REAL SX(1),SY(1)
      INTEGER I,INCX,INCY,IX,IY,M,MP1,N
      IF(N.LE.0) RETURN
      IF(INCX.EQ.1.AND.INCY.EQ.1) GO TO 20
      IX = 1
      IY = 1
      IF(INCX.LT.0)IX = (-N+1)*INCX + 1
      IF(INCY.LT.0)IY = (-N+1)*INCY + 1
      DO 10 I = 1,N
      SY(IY) = SX(IX)
      IX = IX + INCX
      IY = IY + INCY
10    CONTINUE
      RETURN
20    M = MOD(N,7)
      IF(M.EQ.0) GO TO 40
      DO 30 I=1,M
      SY(I) = SX(I)
30    CONTINUE
      IF(N.LT.7) RETURN
40    MP1 = M + 1
      DO 50 I = MP1,N,7
      SY(I) = SX(I)
      SY(I + 1) = SX(I + 1)
      SY(I + 2) = SX(I + 2)
      SY(I + 3) = SX(I + 3)
      SY(I + 4) = SX(I + 4)
      SY(I + 5) = SX(I + 5)
      SY(I + 6) = SX(I + 6)
50    CONTINUE
      RETURN

```

END

C

SUBROUTINE SAXPY(N,SA,SX,INCX,SY,INCY)

REAL SX(1),SY(1),SA

INTEGER I,INCX,INCY,IX,IY,M,MP1,N

IF(N.LE.0) RETURN

IF(SA .EQ. 0.0) RETURN

IF(INCX.EQ.1.AND.INCY.EQ.1) GO TO 20

IX = 1

IY = 1

IF(INCX.LT.0)IX = (-N+1)*INCX + 1

IF(INCY.LT.0)IY = (-N+1)*INCY + 1

DO 10 I = 1,N

SY(IY) = SY(IY) + SA*SX(IX)

IX = IX + INCX

IY = IY + INCY

10 CONTINUE

RETURN

20 M = MOD(N,4)

IF(M.EQ.0) GO TO 40

DO 30 I = 1,M

SY(I) = SY(I) + SA*SX(I)

30 CONTINUE

IF(N.LT.4) RETURN

40 MP1 = M + 1

DO 50 I = MP1,N,4

SY(I) = SY(I) + SA*SX(I)

SY(I + 1) = SY(I + 1) + SA*SX(I + 1)

SY(I + 2) = SY(I + 2) + SA*SX(I + 2)

SY(I + 3) = SY(I + 3) + SA*SX(I + 3)

50 CONTINUE

RETURN

END

C

SUBROUTINE STRDI(T,LDT,N,DET,JOB,INFO)

INTEGER LDT,N,JOB,INFO

REAL T(LDT,1),DET(2)

REAL TEMP

REAL TEN

INTEGER I,J,K,KB,KM1,KP1

IF(JOB/100 .EQ. 0) GO TO 70

DET(1) = 1.0E0

DET(2) = 0.0E0

TEN = 10.0E0

DO 50 I = 1,N

DET(1) = T(I,I)*DET(1)

IF(DET(1) .EQ. 0.0E0) GO TO 60

10 IF(ABS(DET(1))) .GE. 1.0E0) GO TO 20

DET(1) = TEN*DET(1)

DET(2) = DET(2) - 1.0E0

GO TO 10

20 CONTINUE

30 IF(ABS(DET(1))) .LT. TEN) GO TO 40

DET(1) = DET(1)/TEN

DET(2) = DET(2) + 1.0E0

GO TO 30


```

40    CONTINUE
50    CONTINUE
60    CONTINUE
70    CONTINUE
      IF(MOD(JOB/10,10) .EQ. 0) GO TO 170
      IF(MOD(JOB,10) .EQ. 0) GO TO 120
      DO 100 K = 1,N
      INFO = K
      IF(T(K,K) .EQ. 0.0E0) GO TO 110
      T(K,K) = 1.0E0/T(K,K)
      TEMP = -T(K,K)
      CALL SSCAL(K-1,TEMP,T(1,K),1)
      KP1 = K + 1
      IF(N.LT.KP1) GO TO 90
      DO 80 J = KP1,N
      TEMP = T(K,J)
      T(K,J) = 0.0E0
      CALL SAXPY(K,TEMP,T(1,K),1,T(1,J),1)
80    CONTINUE
90    CONTINUE
100   CONTINUE
      INFO = 0
110   CONTINUE
      GO TO 160
120   CONTINUE
      DO 150 KB = 1,N
      K = N + 1 - KB
      INFO = K
      IF(T(K,K) .EQ. 0.0E0) GO TO 180
      T(K,K) = 1.0E0/T(K,K)
      TEMP = -T(K,K)
      IF(K.NE.N) CALL SSCAL(N-K,TEMP,T(K+1,K),1)
      KM1 = K - 1
      IF(KM1 .LT. 1) GO TO 140
      DO 130 J = 1,KM1
      TEMP = T(K,J)
      T(K,J) = 0.0E0
      CALL SAXPY(N-K+1,TEMP,T(K,K),1,T(K,J),1)
130   CONTINUE
140   CONTINUE
150   CONTINUE
      INFO = 0
160   CONTINUE
170   CONTINUE
180   CONTINUE
      RETURN
      END

```

1. Subrotinas principais:

• Subrotina *Solve* estima os parâmetros de um modelo não-linear pelo método de Gauss-Newton. A subrotina *EVAL* é chamada NPTS vezes para construir a matriz de primeiras derivadas parciais baseada nas estimativas iniciais. Um vetor de diferenças é

também construído subtraindo os valores preditos, obtidos da subrotina EVAL, dos valores da variável dependente. A subrotina SQRDC é então usada para decompor a matriz de primeiras derivadas e os resultados desta decomposição QR são usados pela rotina SQRSL para encontrar correções para as estimativas do parâmetro. Este procedimento é repetido com a nova estimativa até que a convergência ocorra ou até que um número especificado de iterações seja excedido. Se a convergência é atingida, a subrotina SOLVE então chama a subrotina DSPLAY para imprimir os resultados.

Subrotinas requeridas:

- Externa: Eval
- Nonlin: DSPLAY
- Linpack: SQRDC, SQRSL, STRDI, SAXPY, SDOT, SSCAL, SSWAP, SNRM2, SCOPY
- Fortran: SIN, COS, ATAN2, ABS, SIGN, MOD, SQRT, AMAX1, MIN0, FLOAT

• Subrotina *Bates* calcula as medidas de curvatura intrínseca e paramétrica de Bates e Watts e a medida do vício dos parâmetros estimados (Box, 1971). Repetidamente chama a subrotina externa EVAL para a construção das matrizes de derivadas de primeira e segunda ordens, baseada nas estimativas dos parâmetros. Cada matriz é padronizada pelo raio padrão e a decomposição QR é aplicada a matriz de primeiras derivadas usando a rotina SQRDC. Os resultados desta decomposição são, então, usados para transformar a matriz de segundas derivadas em porção intrínseca e porção paramétrica da matriz de aceleração, da qual as medidas de curvatura máxima são calculadas pelo procedimento iterativo. A porção paramétrica da matriz de aceleração é também usada para obter as medidas do vício de Box. A subrotina WATTS é então chamada para imprimir os resultados.

Subrotinas requeridas:

- Externa: Eval
- Nonlin: ARRAYS, ACCEL, BOX, CURVE, CNTEST, WATTS
- Linpack: SQRDC, SAXPY, SDOT, SSCAL, SSWAP, SNRM2
- Fortran: SIN, COS, ATAN2, ABS, SIGN, MOD, SQRT, AMAX1, MIN0, FLOAT

2. Subrotinas Auxiliares:

- Subrotina *Accel* transforma a matriz de segundas derivadas em matriz de aceleração usando a decomposição QR da matriz de primeiras derivadas.
- Subrotina *Watts* imprime os resultados obtidos da subrotina Bates.
- Subrotina *Arrays* imprime a matriz de primeiras derivadas (Derivs) e as np faces da matriz de segundas derivadas (A).

- Subrotina *Box* calcula o vício de Box para cada parâmetro estimado e retorna esses valores na matriz de vício.
- Subrotina *Cntest* estipula um teste de convergência para a determinação iterativa das curvaturas intrínseca e paramétrica.
- Subrotina *Curve* calcula as medidas de curvatura intrínseca e paramétrica pelo procedimento iterativo que converge quando o número máximo de iterações é excedido.
- Subrotina *Dsplay* imprime os resultados da subrotina Solve.
- Subrotina *Sort* ordena em ordem crescente de magnitude a parte da matriz real entre A(ii) e A(jj), inclusive.
- Subrotina *Eval* é externa e devemos usá-la para permitir o cálculo dos valores ajustados e o cálculo das primeiras e segundas derivadas parciais do modelo.

A seguir, apresentamos uma tabela com as principais variáveis¹:

<i>Variáveis</i>	<i>Definição</i>
Y	real (k), $k \geq \text{NPTS}$ - Número de observações da var. dependente
Param	real (k) $\geq \text{NP}$ - Estimativas iniciais dos parâmetros
Alpha	real - Critério de Convergência
NITS	inteiro - Número máximo de iterações
NPTS	inteiro - Número de observações
NP	inteiro - Número de parâmetros
Varc	real (ND1,k), $\text{ND1} \geq \text{NP}$, $k \geq \text{NP}$ - matriz de variância e covariância
Derivs	real (k) $\geq \text{NP} * \text{NPTS}$ - inicialmente guarda as derivadas parciais de 1a. ordem e depois os componentes da decomposição <i>QR</i>
Qraux	real (k) $\geq \text{NP}$ - guarda informação da decomposição <i>QR</i>
JPVT	inteiro $\geq \text{NP}$ - arquiva colunas trocadas durante a decomposição <i>QR</i>
Diff	real (k) $\geq \text{NPTS}$ - vetor de diferenças entre val. atuais e ajustados
Wk1	real (k) $\geq \text{NP} * (\text{NP} + 1) / 2$ - vetor de trabalho
Wk2	real (k) $\geq \text{NPTS}$ - vetor temporário de armazenamento
A	real (k) $\geq \text{NPTS} * \text{NP} * \text{NP}$ - inicialmente armazena as derivadas parciais de 2a. ordem e depois a matriz de aceleração
RSS	real - Soma de Quadrado Residual
BIAS	real (k), $k \geq \text{NP}$ - Estimativas do vício de Box

¹Para maiores informações sobre o programa e variáveis dos mesmos, ver Ratkowsky, (1983).

2. Programa para o cálculo da densidade a posteriori marginal para θ_2 , considerando a densidade a priori não informativa, obtida pelo método de Simpson.

```

PROGRAM INTEGRAL_POR_SIMPSON;
uses crt;
TYPE REAL=EXTENDED;

{Variáveis para a função}
var S1, S4, S5, S6, Teta2 : real;
Y,Z:ARRAY[1..12] of real;

PROCEDURE SOMAS (Teta1, Teta2 : real);
Var i : integer;
begin
    S1 := 0; S4 := 0; S5 := 0; S6 := 0;
    for I:=1 to 12 do
        begin
            S1:=S1+sqr(Y[I]-Teta1*Z[I]/(Teta2+Z[I]));
            S4:=S4+sqr(Z[I]/(Teta2+Z[I]));
            S5:=S5+sqr(Z[I])/sqr(sqr(Teta2+Z[I]));
            S6:=S6+sqr(Z[I])/(sqr(Teta2+Z[I])*(Teta2+Z[I]));
        end;
    end;

Function FUNCAO (Teta1, Teta2 : real) : real;
Var G : real;
Const C=1.814000095E+18;
begin
    SOMAS(Teta1, Teta2);
    G := S4*S5-sqr(S6);
    FUNCAO := Teta1*sqrt(G)*exp(-6*ln(S1))/C;
end;

Function f(x : real) : real;
begin
    f := Funcao(x, Teta2);
end;

Function SIMPSON (a, b : real; n : integer) : real;
Var x, h, int : real;
    i : integer;
begin
    h := (b-a)/n;
    int := f(a)+f(b);
    x := a;
    for i := 1 to (n-1) do
        begin

```

```

        x := x+h;
        if odd(i) then
            int:=int + 4*f(x)
        else
            int:=int + 2*f(x);
        end;
        Simpson := int * h/3;
    end;

PROCEDURE PAUSA;
var ch : char;
begin
    while KeyPressed do ch := ReadKey;
    if ReadKey = #0 then
        ch := ReadKey;
end;

Var    INF1, SUP1, INF2, SUP2 : real; {Limites}
        SAIDA, DADOS: text;
        T, P: ARRAY[1..500] of real; {Valores de Saida}
        i, j, n : integer;
        H2 : real;
BEGIN
    clrscr;
    {Lê vetor para a função}
    assign(DADOS,'B:DADOS2.DAT');
    reset(DADOS);
    for I:=1 to 12 do read(DADOS,Z[I],Y[I]);
    close(DADOS);
    {Leitura de dados}
    writeln('DISTRIBUICAO A POSTERIORI MARGINAL P/ TETA2,');
    writeln('OBTIDA PELO METODO DE SIMPSON. ');
    writeln;
    gotoxy(1,6);write('De o limite Inferior para Teta2: ');
    readln(INF2);
    gotoxy(1,8);write('De o limite Superior para Teta2: ');
    readln(SUP2);
    gotoxy(1,10);write('De o valor do incremento: ');
    readln(H2);
    gotoxy(1,15);write('De o limite Inferior para Teta1: ');
    readln(INF1);
    gotoxy(1,17);write('De o limite Superior para Teta1: ');
    readln(SUP1);
    N:=1;
    while odd(N) do
        begin
            gotoxy(1,19);write('De o numero de Divisoões para a integral: (PAR) ');
            readln(N);
        end;

    {Cálculo}

```

```
J:=0; Teta2 := INF2;
while TETA2 ≤ SUP2 do
  begin
    J:= J+1;
    T[J]:= TETA2;
    P[J]:= Simpson (INF1, SUP1, n);
    TETA2:=TETA2+H2;
  end;
clrscr;

{Salva os resultados}
assign(SAIDA,'c:Postt21s.sai');
rewrite(SAIDA);
for I:=1 to J do
  begin
    writeln(T[I]:15:6,' ',P[I]:15:6);
    writeln(SAIDA,T[I]:8:4,P[I]:15:6);
  end;
close(SAIDA);
pausa;

END.
```


Referências e Bibliografia Consultada

- ABRAMOWITZ, M.; STEGUN, I.** (1964) - Handbook of Mathematical Functions, Dover.
- ACHCAR, J. A.** (1990) - "Reparametrization and Accuracy of Laplace. Approximations for Posterior Moments". *Revista de Matemática e Estatística* 8, 23-30.
- ACHCAR, J. A.** (1993) - "Some Aspects of Reparametrization for Statistical Models". *Notas do ICMSC, Série Estatística*, 01.
- ACHCAR, J. A.; SMITH, A. F. M.** (1990) - "Aspects of Reparametrization in Approximate Bayesian Inference". *Bayesian and Likelihood Methods in Statistics and Econometrics*. North-Holland, Elsevier Science Publishers B.V., Amsterdam, North-Holland.
- AITCHISON, J.; DUNSMORE, I. R.** (1975) - Statistical Prediction Analysis. Cambridge: Cambridge University Press.
- AMEMIYA, T.** (1983) - "Nonlinear Regression Models". In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Vol I, 333-389. North-Holland: Amsterdam.
- ATKINSON, A. C.** (1985) - Plots, Transformations and Regressions. Oxford Clarendon Press.
- BATES, D. M.** (1978) - "Curvature Measures of Nonlinearity". Ph.D Thesis, Queen's University at Kingston.
- BATES, D. M.; WATTS, D. G.** (1980) - "Relative Curvature Measures of Nonlinearity" (with discussion). *Journal of the Royal Statistical Society., Series B*, 42, 1-25.
- BATES, D. M.; WATTS, D. G.** (1988) - Nonlinear Regression Analysis and its Applications, John Wiley & Sons.
- BEALE, E. M. L.** (1960) - "Confidence Regions in Nonlinear Estimation". *Journal of the Royal Statistical Society., Series B*, 22, 41-76.

- BICKEL, P. J.; DOKSUM, K. A.** (1977) - Mathematical Statistics. Basic Ideas and Selected Topics. San Francisco, Holden Day.
- BOX, G. E. P.** (1953b) - "A Note on Regions for Test of Kurtosis," *Biometrika* **40**, 465.
- BOX, G. E. P.; COUTIE, G. A.** (1956) - "Application of Digital Computers in the Exploration of Functional Relationships". *Proc. I. E. E.* **103**, Part B, Suppl. 1, 100-107.
- BOX, G. E. P.; COX, D. R.** (1964) - "The Analysis of Transformations" (with discussion). *Journal of the Royal Statistical Society., Series B*, **26**, 211-252.
- BOX, G. E. P.; TIAO, G. C.** (1973) - Bayesian Inference in Statistical Analysis. New York, Addison-Wesley: Reading, Mass.
- BOX, M. J.** (1971) - "Bias in Nonlinear Estimation". *Journal of the Royal Statistical Society., Series B*, **33**, 171-201.
- CONTE, S. D.; DE BOOR, C.** (1965) - Elementary Numerical Analysis: an Algorithmic Approach, McGraw-Hill: New York.
- CURRIE, D. J.** (1982) - "Estimating Michaelis-Menten Parameters: Bias, Variance e Experimental Design", *Biometrics* **38**(4), 907-919.
- DIANANDA, P. H.** (1949) - "Note on Some Properties of Maximum Likelihood Estimates". *Proc. Camb. Phil. Soc.* **45**, 536.
- DONALDSON, J. R.; SCHNABEL, R. B.** (1987) - "Computational Experience with Confidence Region and Confidence Intervals for Nonlinear Least Squares". *Technometrics* **29**, 67-82.
- DONGARRA, J. J.; BUNCH, J. R.; MOLER, C. B.; STEWART, G. W.** (1979) - Linpack Users' Guide, *Society for Industrial and Applied Mathematics*.
- DRAPER, N. R.; SMITH H.** (1981) - Applied Regression Analysis, John Wiley & Sons.

- EAVES, D. M.** (1983) - "On Bayesian Nonlinear Regression with an Enzyme Example", *Biometrika* **70** (2), 373-379.
- EDWARDS, W.; LINDMAN, H.; SAVAGE, L. J.** (1963) - "Bayesian Statistical Inference for Psychological Research", *Psychological Rev.* **70**, 193.
- GALLANT, A. R.** (1975a) - "The Power of the Likelihood Ratio Test of Location in Nonlinear Regression Models". *Journal of the American Statistical Association* **70**, 198-203.
- GALLANT, A. R.** (1975b) - "Nonlinear Regression". *American Statistical* **29** (2), 73-81.
- GALLANT, A. R.** (1987) - Nonlinear Statistical Models, Wiley.
- GELFAND, A.; SMITH, A. F. M.** (1990) - "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association* **85**, 398-409.
- GILL, P. E.; MURRAY, W.; WRIGHT, M. H.** (1981) - Practical Optimization. Academic Press Inc., London.
- GUTTMAN, I.; MEETER, D. A.** (1965) - "On Beale's Measures of Nonlinearity". *Technometrics* **7**, 623-637.
- HAMILTON, D.** (1986) - "Confidence Regions for Parameter Subsets in Nonlinear Regression". *Biometrika* **73**, 57-64.
- HILLS, S. E.; SMITH, A. F. M.** (1993) - "Diagnostics of Posterior Nonnormality in Bayesian Inference", *Biometrika* **80**, 1, 61-74.
- HOUSEHOLDER, A. S.** (1958) - "Unitary Triangularization of a Nonsymmetric Matrix". *Journal of the Association for Computing Machinery* **5**, 339-342.
- JEFFREYS, H.** (1961) - Theory of Probability, (3th ed.) Clarendon: Oxford.
- JENNRICH, R. I.** (1969) - "Asymptotic Properties of Nonlinear Least-Squares Estimations". *The Annals of Mathematical Statistics* **20**, 633-643.

- JENNRICH, R. I.; RALSTON, M. L.** (1979) - "Fitting Nonlinear Models to Data". *Ann. Rev. Biophys. Bioeng.* **8**, 195-238
- KASS, R. E.; SLATE, E. H.** (1992) - Reparametrization and Diagnostics of Posterior Nonnormality in *Bayesian Statistics 4*, Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, 289-306. Oxford University Press.
- KASS, R. E.; SLATE, E. H.** (1994) - "Some Diagnostics of Maximum Likelihood and Posterior Nonnormality", *Annals of Statistics* **22**, 2, 668-695.
- KASS, R. E.; TIERNEY, L.; KADANE, J. B.** (1990) - The Validity of Posterior Expansion based on Laplace's Method, in *Essays in Honor of G. A. Barnard*, ed. J. Hodges, Amsterdam: North-Holland, 473-488.
- KATZ, D.; AZEN, S. P.; SCHUMITZKY, A.** (1981) - "Bayesian Approach to the Analysis of nonlinear models: Implementation and evaluations", *Biometrics* **37**, 137-142.
- LAWLESS, J. F.** (1982) - *Statistical Model and Methods for Lifetime Data*, New York: John Wiley & Sons.
- LINDLEY, D. V.** (1965) - *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Part 2, Inference, Cambridge: Cambridge University Press.
- LINDLEY, D. V.** (1980) - Approximate Bayesian Methods, *Bayesian Statistics*, Proceedings of the First International Meeting Held in Valencia (eds. J. F. M. Smith). Valencia: University Press.
- MARQUARDT, D. W.** (1963) - "An Algorithm for the Estimation of Nonlinear Parameters." *J. SIAM* **11**, 431-441.
- MARQUARDT, D. W.** (1970) - "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation." *Technometrics* **12**, 591.
- MILLER, R. G., Jr.** (1981) - *Simultaneous Statistical Inference*, (2nd ed.). McGraw-Hill: New York.

- MILLIKEN, G. A.; De BRUIN, R. L.; SMITH, J. E. (1977)** - A Simulation study of Hypothesis testing Procedures for the Michaelis-Menten enzyme Kinetic Model. Tech. Report, 36, Dept. of Statistics, Kansas State Univ., Manhattan, Kansas 66506.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. (1974)** - Introductions to the Theory of Statistics, (3th edn). New York, Mcgraw-Hill.
- NAYLOR, J. C.; SMITH, A. F. M. (1982)** - "Applications of a Method for the Efficient Computation of Posterior Distributions", *Applied Statistics* **31**, 214-225.
- PRESS, S. J. (1989)** - Bayesian Statistics: Models and Applications, New York: John Wiley & Sons.
- RAAIJMAKERS, J. G. W. (1987)** - "Statistical Analysis of the Michaelis-Menten equation", *Biometrics* **43**, 793-803.
- RATKOWSKY, D. A. (1983)** - Nonlinear Regression Modelling, Marcel Dekker.
- SAVIN, N. E. (1984)** - Multiple Hypothesis Testing. In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Vol II, 827-879. Elsevier Science Publishers.
- SEBER, G. A. F. (1977)** - Linear Regression Analysis. Wiley: New York.
- SEBER, G. A. F. (1980)** - The Linear Hypothesis; A General Theory. Charles Griffin & Company Ltd., London.
- SEBER, G. A. F.; WILD, C. J. (1989)** - Nonlinear Regression, John Wiley & Sons.
- SNEDECOR, G. W.; COCHRAN, W. G. (1980)** - Statistical Methods, (7th ed.). Ames, Iowa: Iowa State University Press.
- SPROTT, D. A. (1973)** - "Normal Likelihood and their Relation to Large Sample Theory of Estimation. *Biometrika* **60**, 3, 457-465.
- SPROTT, D. A. (1980)** - "Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters". *Biometrika* **67**, 515-523.

- STEWART, G. W.** (1973) - Introduction to Matrix Computations. New York: Academic Press.
- TIERNEY, L.; KADANE, J. B.** (1986) - "Accurate Aproximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association* **81**, 82-86.
- TIERNEY, L.; KASS, R. E.; KADANE, J. B.** (1989) - "Aproximate Marginal Densities for Nonlinear Functions", *Biometrika* **76**, 425-433.
- TRELOAR, M. A.** (1974) - Effects of Puromycin on Galactosyltransferase of Golgi Membranes. Master's Thesis, University of Toronto.
- TURNER, M. C.** (1960) - "On Heuristic Estimation Methods". *Biometrics* **16**, 299.
- WALD, A.** (1943) - "Test of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large", *Trans. Am. Math. Soc.* **54**, 426-482.
- ZELLNER, A.** (1971) - An Introduction to Bayesian Inference in Econometrics. Wiley: New York.