Modelos de regressão sobre dados composicionais

André Pierro de Camargo

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Programa: Matemática Aplicada
Orientador: Prof. Dr. Marcelo de Souza Lauretto

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

São Paulo, outubro de 2011

Modelos de regressão sobre dados composicionais

Esta versão definitiva da dissertação
contém as correções e alterações sugeridas pela
Comissão Julgadora durante a defesa realizada
por André Pierro de Camargo em 9/12/2011.

Comissão Julgadora:

- Prof. Dr. Marcelo de Souza Lauretto (orientador) - EACH-USP
- Profa. Dra. Delhi Paiva Salinas - EACH-USP
- Prof. Dr. Júlio Michael Stern - IME-USP

Agradecimentos

- Agradeço primeiramente aos meus pais por terem me proporcionado um ambiente familiar favorável ao meu desenvolvimento intelectual.

- Agradeço aos meus amigos pelos momentos de descontração necessários para o relaxamento da mente.

- Agradeço aos meus professores pelo conhecimento adquirido direta ou indiretamente, em especial meu orientador (Marcelo Lauretto), o professor Júlio Michael Stern e meus ex-orientadores de iniciação científica: Eduardo do Nascimento Marcos e Paulo Agozzini Martin.

- Agradeço à minha esposa, Roberta, e à minha pequena filha, Marina, pelo apoio e carinho recebidos nesse período.

- Agraddeço a CAPES pelo auxílio financeiro.

Resumo

Dados composicionais são constituídos por vetores cujas componentes representam as proporções de algum montante, isto é: vetores com entradas positivas cuja soma é igual a 1. Em diversas áreas do conhecimento, o problema de estimar as partes $y_1, y_2, \ldots, y_D$ correspondentes aos setores $SE_1, SE_2, \ldots, SE_D$, de uma certa quantidade Q, aparece com frequência. As porcentagens $y_1, y_2, \ldots, y_D$ de intenção de votos correspondentes aos candidatos $Ca_1, Ca_2, \ldots, Ca_D$ em eleições governamentais ou as parcelas de mercado correspondentes a industrias concorrentes formam exemplos típicos. Naturalmente, é de grande interesse analisar como variam tais proporções em função de certas mudanças contextuais, por exemplo, a localização geográfica ou o tempo. Em qualquer ambiente competitivo, informações sobre esse comportamento são de grande auxílio para a elaboração das estratégias dos concorrentes.

Neste trabalho, apresentamos e discutimos algumas abordagens propostas na literatura para regressão sobre dados composicionais, assim como alguns métodos de seleção de modelos baseados em inferência bayesiana.

Palavras-chave: Modelos de regressão, Dados composicionais, Seleção de modelos, BIC, FBST.

Abstract

Compositional data consist of vectors whose components are the proportions of some whole. The problem of estimating the portions $y_1, y_2, \ldots, y_D$ corresponding to the pieces $SE_1, SE_2, \ldots, SE_D$ of some whole Q is often required in several domains of knowledge. The percentages $y_1, y_2, \ldots, y_D$ of votes corresponding to the competitors $Ca_1, Ca_2, \ldots, Ca_D$ in governmental elections or market share problems are typical examples. Of course, it is of great interest to study the behavior of such proportions according to some contextual transitions. In any competitive environmet, additional information of such behavior can be very helpful for the strategists to make proper decisions.

In this work we present and discuss some approaches proposed by different authors for compositional data regression as well as some model selection methods based on bayesian inference.

Keywords: Regression models, Compositional data, Model selection, BIC, FBST.

Sumário

Lista de Abreviaturas **ix**

Lista de Símbolos **xi**

Lista de Figuras **xiii**

Lista de Tabelas **xv**

1 Introdução **1**

1.1 Preliminares 3

2 Modelos de regressão de Dirichlet **5**

2.1 A distribuição de Dirichlet 5

2.2 Regressão de Dirichlet linear 6

2.3 Estimação de parâmetros 7

2.3.1 Algoritmo de Hijazi-Jernigan para seleção de valores iniciais 8

2.3.2 Novo algoritmo para seleção de valores iniciais 8

2.3.3 Resultados numéricos 9

2.4 Eliminação das restrições sobre o espaço paramétrico 10

3 Outros modelos de regressão **13**

3.1 Transformações $S_D \longmapsto \mathbb{R}^{D-1}$ 13

3.1.1 Abrangência do método 14

3.2 Coordenadas esféricas 18

3.3 Outras transformações no Simplex 20

3.3.1 Observações 23

3.4 Método dos mínimos quadrados no Simplex 23

4 Aplicações **25**

4.1 Lago Ártico 25

4.1.1 Modelo de Dirichlet Linear 25

4.1.2 Modelo de Melo, Vasconcellos e Lemonte 26

4.1.3 Modelo linear Logaritmo da Razão/ Log-esférico 27

4.1.4 Modelo linear Tangente-esférico 28

4.2 Despesas domésticas (bivariado) 29

4.2.1 Modelo de Dirichlet Linear 29

4.2.2 Modelo de Melo, Vasconcellos e Lemonte 30
4.2.3 Modelo linear $TgRatio$ (3.22) 31
4.3 Casamentos por faixa etária (bivariado) 33
4.3.1 Modelo de Dirichlet Linear 33
4.3.2 Modelo de Melo, Vasconcellos e Lemonte 34
4.3.3 Modelo linar Log-esférico 35
4.4 Considerações sobre o Espaço Paramétrico 36
4.4.1 Modelo de Dirichlet/ Mínimos quadrados (linear) 36

5 Seleção de modelos **39**
5.1 BIC 39
5.1.1 Uma proposta de formalização do BIC 42
5.2 FBST 45
5.2.1 Definição formal 45
5.3 Teste da Razão de verossimilhança (LR) 46
5.4 FBST vs LR 46
5.5 Aplicações do BIC aos modelos de Dirichlet linear 47
5.5.1 Lago Ártico 47
5.5.2 Despesas domésticas 48

6 Conclusões **51**

A Cálculo dos coeficientes da regressão linear por mínimos quadrados **53**

B Base de dados **55**
B.1 Lago Ártico 56
B.2 Despesas Domésticas 57
B.3 Casamentos por faixa-etária 58

Referências Bibliográficas **59**

Índice Remissivo **61**

Lista de Abreviaturas

BIC	Critério de Informação Bayesiano (*Bayesian information criterion*).
FBST	Teste de significância integralmente bayesiano (*Full Bayesian Significant Test*).
LR	Teste da Razão de verossimilhança (*Likelihood-ratio test*).

Lista de Símbolos

$\mathbb{R}^q$	Espaço euclidiano q-dimensional.
S_D	Simplex (elementos com D componentes).
d	$d = D - 1$.
$z_\bullet$	Covariável.
$x_\bullet$	Covariável estendida.
$X = (x_{ij})$	Matriz da covariável estendida (definida na introdução).
N^d	Distribuição normal d-variada.
Γ	Função Gama.
Dir	Distribuição de Dirichlet.
E	Valor esperado/ Esperança.
β	Matriz de parâmetros.
$Prev$	Função de previsão (definida na introdução).
ψ	Função de parâmetros (definida na introdução).
L	Função de verossimilhança.
δ	Notação delta de Kronecker.
∇	Gradiente.
Δ	Distância de Aitchison.
$\|\| \ \|\|$	Norma Euclidiana.
M_d	Transformação Logística multiplicativa.
H_d	Transformação Logística híbrida.
alr	Transformação Logaritmo da razão assimétrica.
clr	Transformação Logaritmo da razão centrada.
M_{mod}	Matriz de modelo.
Θ	Espaço paramétrico.

Lista de Figuras

2.1 Comparativo entre os métodos de seleção de valores iniciais para o modelo de Dirichlet linear . . . 10

3.1 Interpretação geométrica da transformação *alr* . . . 20
3.2 Parametrização do 3-Simplex por coordenadas esféricas . . . 21

4.1 Modelos de Dirichlet lineares . . . 26
4.2 Modelos de Melo, Vasconcellos e Lemonte. . . . 26
4.3 Modelo linear Log-esférico vs Modelo linear Logaritmo da razão . . . 28
4.4 Modelos Tangente-esféricos . . . 29
4.5 Modelo de Dirichlet linear . . . 30
4.6 Modelo de Melo, Vasconcellos e Lemonte . . . 31
4.7 Modelo linear *TgRatio* . . . 32
4.8 Modelo de Dirichlet Linear . . . 33
4.9 Modelo de Melo, Vasconcellos e Lemonte . . . 34
4.10 Modelo Log-esférico . . . 36
4.11 Sensibilidade da verossimilhança de MDQ sobre dilatações . . . 37
4.12 Modelos quadráticos restritos à casca esférica unitária . . . 38

5.1 (a) Erro Tipo I, (b) Erro Tipo II, (c) Erro médio e (d) Erro Tipo II empírico. . . . 47
5.2 Submodelos de MDQ . . . 48
5.3 Melhor modelo segundo Schwarz . . . 49

Lista de Tabelas

3.1 Transformações logísticas elementares $y \in S_D \longmapsto v \in \mathbb{R}^d$ 15

4.1 Parâmetros de máxima verossimilhança 26
4.2 Parâmetros de máxima verossimilhança 27
4.3 Parâmetros de máxima verossimilhança 28
4.4 Parâmetros de máxima verossimilhança 29
4.5 Parâmetros de máxima verossimilhança 30
4.6 Parâmetros de máxima verossimilhança 31
4.7 Parâmetros de máxima verossimilhança 32
4.8 Parâmetros de máxima verossimilhança 34
4.9 Parâmetros de máxima verossimilhança 35
4.10 Parâmetros de máxima verossimilhança 35
4.11 Transformação de MDQ para coordenadas esféricas 38
4.12 Parâmetros de máxima verossimilhança/ Mínimos quadrados 38

5.1 Alguns submodelos de MDQ ordenados do menor para o maior BIC 48

B.1 Composição do solo do Lago Ártico em função da profundidade. 56
B.2 Despesas domésticas: T = Total gasto (em HK$); A = Alimentação; O = Outros Bens; S = Serviços. 57
B.3 Porcentagen de casamentos por faixa etária e sexo. Fonte: IBGE 58

Capítulo 1

Introdução

Em diversas áreas do conhecimento, o problema de estimar as partes $y_1, y_2, \ldots, y_D$ correspondentes aos setores $SE_1, SE_2, \ldots, SE_D$, de uma certa quantidade Q, aparece com frequência. As porcentagens $y_1, y_2, \ldots, y_D$ de intenção de votos correspondentes aos candidatos $Ca_1, Ca_2, \ldots, Ca_D$ em eleições governamentais ou as parcelas de mercado correspondentes a industrias concorrentes formam exemplos típicos. Naturalmente, é de grande interesse analisar como variam tais proporções em função de certas mudanças contextuais, por exemplo, a localização geográfica ou o tempo. Em qualquer ambiente competitivo, informações sobre esse comportamento são de grande auxílio para a elaboração das estratégias dos competidores.

Definição: Chamamos de um *dado composicional* a cada D-upla $(y_1, y_2, \ldots, y_D)$ de números positivos tais que $\sum_{j=1}^{D} y_j = 1$. O conjunto

$$S_D = \{(y_1, y_2, \ldots, y_D) \in \mathbb{R}^D \ : \ \sum_{j=1}^{D} y_j \ = \ 1, \ y_j > 0, \ j = 1, 2, \ldots, D\}$$

é denominado o $D-1$-Simplex.

Observado um conjunto de dados $\{(z_{1\bullet}, y_{1\bullet}), (z_{2\bullet}, y_{2\bullet}), \ldots (z_{n\bullet}, y_{n\bullet})\}$, onde $y_{j\bullet} \in S_D$ e $z_{j\bullet} \in \Omega \subseteq \mathbb{R}^C$, $j = 1, 2, \ldots, n$, deseja-se obter um modelo de regressão paramétrico $y_\bullet \sim z_\bullet$. O sucesso dessa tarefa depende, essencialmente, de dois fatores: da escolha da família paramétrica de modelos e do método de inferência utilizado. Um método bastante utilizado consiste em:

1. Selecionar uma família paramétrica de modelos probabilísticos $G = \{g(y_\bullet|\psi) : \psi \in \Psi\}$ (aqui descritos por suas funções densidade de probabilidade), definidos em S_D .

2. Supor que, para cada valor de $z_\bullet$, $y_\bullet|z_\bullet$ é distribuída segundo G com parâmetro $\psi(z_\bullet, \beta)$, cuja dependencia com $z_\bullet$ é calibrada por um vetor (ou matriz) β.

3. Ajustar o vetor/matriz de parâmetros β.

4. Estimar o valor de $y_\bullet$ em $z_\bullet$, $Prev(z_\bullet)$, como o valor esperado da distribuição G com parâmetro $\psi(z_\bullet, \beta)$.

Assumindo que as variáveis $y_{j\bullet}|z_{j\bullet}, j = 1, 2, \ldots, n$ são independentes, podemos formar a função de verossimilhança com respeito ao vetor/matriz de parâmetros β para estimar seus valores via máxima verossimilhança, mínimos quadrados ou outro método de inferência.

Neste trabalho investigamos e estendemos algumas abordagens propostas na literatura baseadas no método descrito acima. O texto está assim organizado:

No segundo Capítulo, apresentamos dois modelos de regressão descritos em Campbell e Mosimann (1987) e Melo *et al.* (2009), ambos baseados na distribuição de Dirichlet (a distribuição mais conhecida definida no Simplex, Pereira e Stern (2008)). A nossa principal contribuição referente ao conteúdo desse Capítulo é o desenvolvimento de um algoritmo mais eficiente para encontrar um ponto inicial viável para ser fornecido aos algoritmos de maximização da função de verossimilhança (Seção 2.3.2).

No terceiro Capítulo, apresentamos os modelos de regressão descritos nos trabalhos de Aitchinson (1986) e Wang *et al.* (2007). O primeiro propõe uma alternativa à utilização de distribuições no Simplex, transformando os dados composicionais para o Espaço Euclidiano, modelando os dados transformados e retornando para o Simplex. O segundo descreve os dados composicionais em termos de coordenadas esféricas generalizadas e modela os ângulos diretamente em função da covariável X. A nossa principal contribuição referente ao conteúdo desse Capítulo consiste na completa caracterização das transformações equivalentes (Seção 3.1.1) e na elaboração de novas transformações entre o Simplex e o Espaço Euclidiano (Seção 3.3).

No quarto Capítulo, aplicamos os modelos de regressão descritos nos Capítulos 2 e 3 a três conjuntos de dados (disponíveis para consulta no Apêndice B):

1. O primeiro é referente a 39 composições do solo do Lago Ártico (areia, lodo e argila em função da profundidade) apresentados por Coakley e Rust Coakley e Rust (1968) e adaptados por Aitchinson (1986). A análise do solo, em geral, é um fator discriminante na área da construção civil.

2. O segundo é referente ao conjunto de dados encontrado em Aitchinson (1986), constituído pelas despesas domésticas (separadas em Manutenção domiciliar, Alimentação, Outros bens e Serviços.) de 20 homens e 20 mulheres, em função do total gasto.

3. O terceiro corresponde a um conjunto de dados fornecido pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e disponível em *http://seriesestatisticas.ibge.gov.br/* relativo à variação das porcentagens de casamentos por faixa etária e sexo no período 1984-2002.

   A análise desses outros dois conjuntos de dados é de grande interesse para as ciências sociais. O segundo constitui uma típica pesquisa mercadológica e o estudo do terceiro pode auxiliar na descrição de outros fenômenos direta ou indiretamente correlacionados, como, por exemplo, a disseminação de doenças sexualmente transmissíveis.

No quinto Capítulo expomos três métodos de seleção de modelos baseados em inferência bayesiana: BIC, FBST e Teste da razão de verossimilhança. Nossa principal contribuição referente ao conteúdo desse Capítulo é um estudo detalhado do critério de Schwarz (BIC).

1.1 Preliminares

Ao longo do texto, as composições observadas serão denotadas por $y_{j\bullet} = (y_{j1}, y_{j2}, \ldots, y_{jn}), j = 1, 2, \ldots, n$ e $Y = (y_{ij})$ denota a matriz das composições. Um dado composicional genérico (ou variável aleatória assumindo valores no $D-1$-Simplex) será denotado por y ou $y_\bullet$.

Seguindo a notação padrão em modelos de regressão, em lugar da covariável original $z_\bullet \in \Omega$, trabalharemos com a variável estendida $x_\bullet = (f_1(z_\bullet), f_2(z_\bullet), \ldots, f_k(z_\bullet)) \in \mathbb{R}^k$, onde $f_1, f_2, \ldots, f_k : \Omega \longrightarrow \mathbb{R}$ são transformações apropriadas. Para garantir a unicidade de alguns estimadores ótimos para a matriz β, iremos assumir que a matriz

$$X = (x_{ij}) = \begin{bmatrix} x_{1\bullet} \\ x_{2\bullet} \\ \vdots \\ x_{n\bullet} \end{bmatrix} = \begin{bmatrix} f_1(z_{1\bullet}) & f_2(z_{1\bullet}) & \ldots & f_k(z_{1\bullet}) \\ f_1(z_{2\bullet}) & f_2(z_{2\bullet}) & \ldots & f_k(z_{2\bullet}) \\ \vdots & \vdots & & \vdots \\ f_1(z_{n\bullet}) & f_2(z_{n\bullet}) & \ldots & f_k(z_{n\bullet}) \end{bmatrix} \tag{1.1}$$

possui posto $k \leq n$ (ver apêndice A).

Para simplificar a notação, vetores do Espaço Euclidiano $\mathbb{R}^q$, para q genérico, serão interpretados, eventualmente, como matrizes colunas (quando couber).

Ao longo do texto, $e - s$ e $e + s$ denotarão, respectivamente 10^{-s} e 10^s.

Capítulo 2

Modelos de regressão de Dirichlet

2.1 A distribuição de Dirichlet

O modelo probabilístico de Dirichlet , $Dir(\lambda_1, \lambda_2, \ldots, \lambda_D)$, definido no $D-1$-Simplex, possui função densidade de probabilidade dada por:

$$f(y|\lambda) = \frac{\Gamma(\Lambda)}{\prod_{i=1}^{D} \Gamma(\lambda_i)} \prod_{i=1}^{D} y_i^{\lambda_i - 1}, \tag{2.1}$$

onde $\Lambda = \sum_{i=1}^{D} \lambda_i$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_D) > 0$, $y = (y_1, y_2, \ldots, y_D) \in S_D$ e Γ representa a função Gama: $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$. Fazendo

$$T(y) = (\log(y_1), \log(y_2), \ldots, \log(y_D)), \quad A(\lambda - 1) = \sum_{i=1}^{D} \log(\Gamma(\lambda_i)) - \log(\Gamma(\Lambda)),$$

podemos reescrever (2.1) como $f(y|\alpha) = \exp\left\{ T(y) \times \alpha^t - A(\alpha) \right\}$, $\alpha = (\lambda_1 - 1, \lambda_2 - 1, \ldots, \lambda_D - 1)$,

donde segue que Dir pertence à Familia Exponencial.

Dada uma amostra aleatória $y_{1\bullet}, y_{2\bullet}, \ldots, y_{n\bullet}$ de uma variável $y \sim Dir(\lambda_1, \lambda_2, \ldots, \lambda_D)$, a função de máxima verossimilhança é dada por

$$L(\lambda|y_{1\bullet}, y_{2\bullet}, \ldots, y_{n\bullet}) = \exp\left\{ \left[\sum_{j=1}^{n} T(Y_j) \right] \times \alpha^t - n.A(\alpha) \right\}.$$

A demonstração da existência e unicidade do estimador de máxima verrosimilhança para a Distribuição de Dirichlet pode ser encontradas em Ronning (1989). Demais propriedades da distribuição de Dirichlet podem ser encontradas em Pereira e Stern (2008).

2.2 Regressão de Dirichlet linear

No modelo de regressão de Dirichlet linear, supõe-se que $y_\bullet|x_\bullet$ possui distribuição de Dirichlet com parâmetros dados por

$$\psi(x_\bullet,\beta) = \begin{bmatrix} \lambda_1(x_\bullet) \\ \lambda_2(x_\bullet) \\ \vdots \\ \lambda_D(x_\bullet) \end{bmatrix} = \begin{bmatrix} m_{11}\beta_{11} & m_{12}\beta_{12} & \dots & m_{1k}\beta_{1k} \\ m_{21}\beta_{21} & m_{22}\beta_{22} & \dots & m_{2k}\beta_{2k} \\ \vdots & \vdots & & \vdots \\ m_{D1}\beta_{D1} & m_{D2}\beta_{D2} & \dots & m_{Dk}\beta_{Dk} \end{bmatrix} \times \; x_\bullet \;, \tag{2.2}$$

onde a matriz de modelo $M_{mod} = (m_{ij})$ indica quais dos parâmetros β_{ij} fazem parte do modelo, isto é:

$$\begin{cases} m_{ij} = 1, & \text{se } \beta_{ij} \text{ pertence ao modelo.} \\ m_{ij} = 0, & \text{caso contrário.} \end{cases} \tag{2.3}$$

A matriz de modelo possui um papel importante na identificação das variáveis relevantes do modelo (ver Capítulo 5). Neste Capítulo, porém, vamos assumir (a menos que seja mencionado) que a matriz de modelo é completa, isto é: $m_{ij} = 1 \; \forall \; i,j$. Vale ressaltar que tais modelos possuem algumas propriedades especiais (ver apêndice A). Assim, podemos reescrever (2.2) na forma compacta

$$\psi(x_\bullet,\beta) = \beta \times x_\bullet. \tag{2.4}$$

Campbell e Mosimann (1987) sugeriram um modelo de regressão fazendo os parâmetros $\lambda_1(x_\bullet), \lambda_2(x_\bullet), \dots, \lambda_D(x_\bullet)$ dependerem polinomialmente da covariável $z_\bullet$ (supondo $z_\bullet \in \mathbb{R}$), isto é: $x_\bullet = (1, z_\bullet, \dots, z_\bullet^{k-1})$.

Devido às restrições sobre os parâmetros da distribuição de Dirichlet (devem ser todos positivos), devemos considerar apenas parâmetros β_{ij} e funções $f_1, f_2, \dots, f_k$ tais que $\psi(x_\bullet|\beta) > 0$ para todo valor da covariável estendida $x_\bullet$. Sendo $x_\bullet$ uma variável contínua, essa condição é inviável de ser verificada na prática para uma dada matriz β. Assim, vamos impor somente que

$$\psi(x_{j\bullet},\beta) \; > \; 0 \; \forall \; j \; \in \; \{1,2,\dots,n\}. \tag{2.5}$$

Espera-se[1], assim, que $\psi(x_{j\bullet}|\beta)$ seja positivo para todo valor da covariável extendida $x_\bullet$ (embora

[1] A condição (2.5) garante que $\beta \times x_\bullet > 0$ sempre que $x_\bullet$ for uma combinação linear de $(x_{1\bullet}), (x_{2\bullet}), \dots, (x_{n\bullet})$ com coeficientes positivos.

isso não seja uma garantia).

Estimados os parâmetros β, podemos prever o valor do dado composicional referente ao valor $x_{0\bullet}$ da covariável $x_\bullet$ tomando-se como estimativa o valor esperado da distribuição de Dirichlet com parâmetros $\lambda_1(x_{0\bullet}), \lambda_2(x_{0\bullet}), \ldots, \lambda_D(x_{0\bullet})$, isto é:

$$Prev(x_{0\bullet}) = E(y(x_{0\bullet})), \qquad y(x_{0\bullet}) \sim Dir(\lambda_1(x_{0\bullet}), \lambda_2(x_{0\bullet}), \ldots, \lambda_D(x_{0\bullet})). \tag{2.6}$$

Pelo Teorema dos momentos, ver Pereira e Stern (2008), obtemos:

$$Prev(x_{0\bullet}) = \left(\frac{\lambda_1(x_{0\bullet})}{\Lambda(x_{0\bullet})}, \frac{\lambda_2(x_{0\bullet})}{\Lambda(x_{0\bullet})}, \ldots, \frac{\lambda_D(x_{0\bullet})}{\Lambda(x_{0\bullet})}\right), \qquad \Lambda(x_{0\bullet}) = \sum_{j=1}^{D} \lambda_j(x_{0\bullet}). \tag{2.7}$$

2.3 Estimação de parâmetros

Em seu estudo sobre o trabalho de Campbell e Mosimann, Hijazi e Jernigan (2009) propõem estimar os parâmetros β_{ij} pela maximização da função de verossimilhança:

$$L(\beta|Y_1, Y_2, \ldots Y_n) = \prod_{l=1}^{n} \left[\Gamma(\Lambda(x_{l\bullet})) \prod_{i=1}^{D} \frac{y_{li}^{\lambda_i(x_{l\bullet})-1}}{\Gamma(\lambda_i(x_{l\bullet}))}\right].$$

Assim,

$$\log L \;=\; \sum_{l=1}^{n} \left[\log\Gamma(\Lambda(x_{l\bullet})) \;+\; \sum_{i=1}^{D}(\lambda_i(x_{l\bullet}) - 1) * \log y_{li} - \log\Gamma(\lambda_i(x_{l\bullet}))\right] \tag{2.8}$$

$$\frac{\partial \log L}{\partial \beta_{ij}} = \sum_{l=1}^{n} \left[diGamma\,(\sum_{p=1}^{D} \lambda_p(x_{l\bullet})\,) - \;diGamma\,(\lambda_i(x_{l\bullet})) + \log y_{li}\right] * x_{lj} \tag{2.9}$$

$$\frac{\partial \log L}{\partial \beta_{ab}\partial \beta_{ij}} = \sum_{l=1}^{n} \left[triGamma\,(\sum_{p=1}^{D} \lambda_p(x_{l\bullet})\,) - \;\delta_{ai} * triGamma\,(\lambda_i(x_{l\bullet}))\right] * x_{lj}x_{lb} \tag{2.10}$$

onde $diGamma(u) = \frac{\partial \log\Gamma}{\partial u}(u)$, $triGamma(u) = \frac{\partial^2 \log\Gamma}{\partial u^2}(u)$ e $\delta_{ai} = \begin{cases} 1, & \text{se } a = i \\ 0, & \text{caso contrário} \end{cases}$.

Evidentemente, a complexidade da função $\nabla \log L$ não permite encontrar a solução da equação $\nabla \log L = 0$ explicitamente. Hijazi e Jernigan propuseram um algoritmo para selecionar um valor inicial para algum método iterativo de maximização da função $\log L$, condicionado a (2.5).

2.3.1 Algoritmo de Hijazi-Jernigan para seleção de valores iniciais

1. Extrair r amostras $A_1, A_2, \ldots, A_r$ (todas de tamanho $m, m < n$) do conjunto $\{Z_1, Z_2, \ldots Z_n\}$, $Z_j = (x_{j\bullet}, y_{j\bullet}), j = 1, 2, \ldots, n$.

2. Para cada amostra $A_i = \{Z_{i_1}, Z_{i_2}, \ldots Z_{i_m}\}$, ajuste $\{y_{i_1\bullet}, y_{i_2\bullet}, \ldots, y_{i_m\bullet}\}$ por uma Dirichlet (pelo método da máxima verossimilhança, via algum algoritmo iterativo de maximização) com parâmetros $\widehat{\lambda_{i1}}, \widehat{\lambda_{i2}}, \ldots, \widehat{\lambda_{iD}}$, utilizando as estimativas obtidas pelo método dos momentos (ver Pereira e Stern (2008)) como ponto inicial. Calcule a média amostral da covariavel $z_\bullet$ em cada amostra, isto é: $\overline{z_{i\bullet}} = \frac{1}{m}\sum_{j=1}^{m} z_{i_j\bullet}$ e calcule os valores correspondentes da variável extendida $\overline{x_{i\bullet}}$.

   Obtem-se, assim, as matrizes:

$$\begin{bmatrix} \widehat{\lambda_{11}} & \widehat{\lambda_{12}} & \ldots & \widehat{\lambda_{1D}} \\ \widehat{\lambda_{21}} & \widehat{\lambda_{22}} & \ldots & \widehat{\lambda_{2D}} \\ \vdots & \vdots & & \vdots \\ \widehat{\lambda_{r1}} & \widehat{\lambda_{r2}} & \ldots & \widehat{\lambda_{rD}} \end{bmatrix}, \quad \begin{bmatrix} \overline{x_{1\bullet}} \\ \overline{x_{2\bullet}} \\ \vdots \\ \overline{x_{r\bullet}} \end{bmatrix}.$$

3. Obter os parâmetros β_{ij} que minimizam a norma euclidiana da diferença

$$\begin{bmatrix} \widehat{\lambda_{11}} & \widehat{\lambda_{21}} & \ldots & \widehat{\lambda_{r1}} \\ \widehat{\lambda_{12}} & \widehat{\lambda_{22}} & \ldots & \widehat{\lambda_{r2}} \\ \vdots & \vdots & & \vdots \\ \widehat{\lambda_{1D}} & \widehat{\lambda_{2D}} & \ldots & \widehat{\lambda_{rD}} \end{bmatrix} - \begin{bmatrix} m_{11}\beta_{11} & m_{12}\beta_{12} & \ldots & m_{1k}\beta_{1k} \\ m_{21}\beta_{21} & m_{22}\beta_{22} & \ldots & m_{2k}\beta_{1k} \\ \vdots & \vdots & & \vdots \\ m_{D1}\beta_{D1} & m_{D2}\beta_{D2} & \ldots & m_{Dk}\beta_{Dk} \end{bmatrix} \times \begin{bmatrix} \overline{x_{1\bullet}} & \overline{x_{2\bullet}} & \ldots & \overline{x_{r\bullet}} \end{bmatrix}.$$

4. Utilizar os parâmetros obtidos como valores iniciais.

Vale ressaltar, porém, que o método iterativo de maximização da verossimilhança restrito à condição (2.5) requer que o ponto inicial também a satisfaça e não há garantias de que o algoritmo de Hijazi-Jernigan produza sempre pontos inciais com essa propriedade. Nós desenvolvemos um algoritmo mais eficiente para a obtenção de um ponto inicial.

2.3.2 Novo algoritmo para seleção de valores iniciais

1º caso: O intercepto (função constante igual a 1) pertence ao conjunto das funções $\{f_1, f_2, \ldots, f_k\}$ definidas na Seção 1.1.

1. Suponha que f_1 é a função constante igual a 1, isto é $x_\bullet = (1, f_2(z_\bullet), \ldots)$

2. Introduza parâmetros artificiais $\beta_{j_1 1}, \beta_{j_2 1}, \ldots, \beta_{j_l 1}$, de modo a completar a primeira coluna da matriz de modelo, caso seja necessário. Para $D = 4, k = 3$ e

$$\begin{bmatrix} \lambda_1(x_\bullet) \\ \lambda_2(x_\bullet) \\ \lambda_3(x_\bullet) \\ \lambda_4(x_\bullet) \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} & \beta_{13} \\ \beta_{21} & 0 & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \\ 0 & \beta_{42} & \beta_{43} \end{bmatrix} \times \; x_\bullet \quad ,$$

por exemplo, devemos completar com β_{11} e $\beta_{4,1}$, isto é: $l = 2, j_1 = 1$ e $j_2 = 4$.

3. Ajuste os dados observados $y_{1\bullet}, y_{2\bullet}, \ldots y_{n\bullet}$ por uma Dirichlet com parâmetros $\beta_{11}, \beta_{21}, \ldots, \beta_{D1}$.

O ponto incial $\beta^* = \begin{bmatrix} \beta_{11} & \beta_{12} = 0 & \beta_{13} = 0 & \ldots & \beta_{1k} = 0 \\ \beta_{21} & \beta_{22} = 0 & \beta_{23} = 0 & \ldots & \beta_{1k} = 0 \\ \vdots & \vdots & & \vdots & \\ \beta_{D1} & \beta_{D2} = 0 & \beta_{D3} = 0 & \ldots & \beta_{Dk} = 0 \end{bmatrix}$ certamente satisfaz (2.5).

4. Obtenha o ponto de máximo, por maximazação restrita à (2.5), da função

$\widetilde{L} = \log L(\beta | y_{1\bullet}, y_{2\bullet}, \ldots y_{n\bullet}) - M(\eta) \sum_{r=1}^{l} (\beta_{j_r 1})^2$, utilizando o ponto inicial β^*, onde $M(\eta)$ é uma quantidade positiva que começa com um valor baixo e vai aumentando ao longo do número de iterações η.

5. Descarte os valores correspondentes aos parâmetros artificiais adicionados (os valores descartados deverão ser próximos de zero).

2º caso: O intercepto não pertence ao conjunto das funções $\{f_1, f_2, \ldots, f_k\}$.

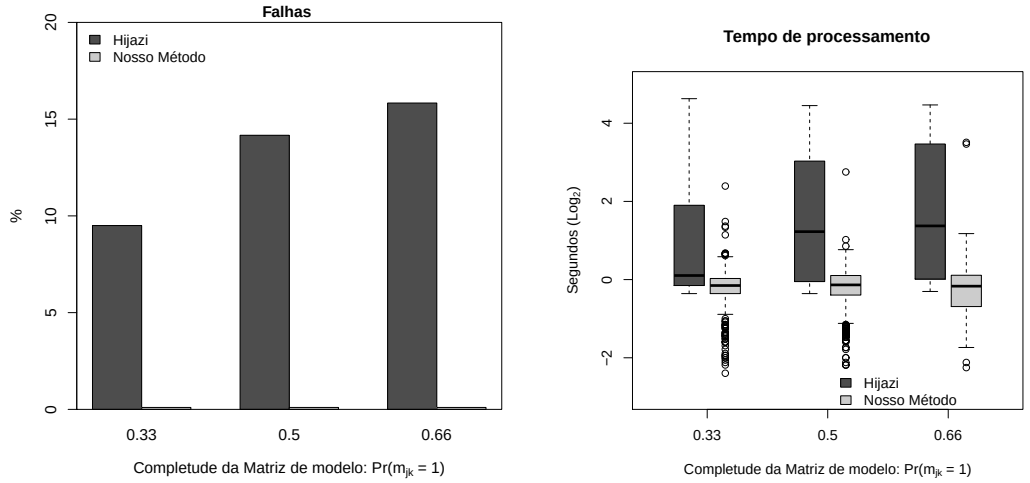
1. Repita o 1º caso adicionando a função constante igual a 1 à esse conjunto (note que, nesse caso, serão introduzidas D variáveis artificiais)

Observe que a função $\widetilde{L}$ penaliza os parâmetros eventualmente adicionados de tal forma que, ao fim do processo de maximização iterativo, os valores dos parâmetros adicionados deverão ser próximos a zero. Além disso, o fato de β^* sempre satisfazer (2.5) garante um método mais estável de obtenção dos parâmetros de máxima verossimilhança.

2.3.3 Resultados numéricos

Para avaliar a eficiência dos métodos propostos para selecionar um valor inicial no modelo de Dirichlet linear, realizamos algumas simulações baseadas em modelos polinomiais de grau menor ou igual a 3 ($k = 4,\ z_\bullet \in \mathbb{R},\ x_\bullet = (1,\ z_\bullet,\ z_\bullet^2,\ z_\bullet^3)$ com um conjunto de subamostras do conjunto de dados do Lago Ártico. Foram selecionadas (aleatoriamente) m =1000 subamostras com 50% (n = 20) e 70% (n = 27) dos pontos do conjunto de dados original. Para cada subamostra, foram definidas, por processos de Bernoulli, três Matrizes de modelo com porcentagens de 1's próximas a $p = 0.33, p = 0.5$

e $p = 0.66$. Para evitar modelos inconsistentes, também foi imposto que tais matrizes deveriam apresentar ao menos uma entrada não nula por linha. Duas medidas de eficiência foram analisadas: (1) A porcentagen dos casos em que os algoritmos não foram capazes de encontrar pontos iniciais viáveis (limitado a trinta tentativas[2] por subamostra); (2) O tempo médio de processamento de cada método. Os resultados são mostrados na Figura 2.1. O tamanho das subamostras (20 ou 27) pareceu não alterar a performance e, assim, os resultados são mostrados juntos.



(a) (b)

Figura 2.1: *Comparativo entre os métodos de seleção de valores iniciais para o modelo de Dirichlet linear*

O gráfico da esquerda mostra claramente uma maior estabilidade do nosso algoritmo: em todos os 6000 casos conseguimos um ponto inicial dentro da região viável. É claro que o tempo de processamento depende da forma como os algoritmos foram implementados. Com a nossa implementação, o nosso algoritmo mostrou também uma superioridade com relação ao outro, mostrada no gráfico da direita: a distribuição dos tempos nos três casos, para o nosso algoritmo, está mais concentrada em torno das medianas que, por sua vez, são inferiores às correspondentes produzidas pelo algoritmo de Hijazi-Jernigan.

2.4 Eliminação das restrições sobre o espaço paramétrico

Uma variante do modelo de regressão linear, proposto em Melo *et al.* (2009), consiste em escolher D funções positivas $g_j : \mathbb{R} \longrightarrow \mathbb{R}^+,\ j = 1, 2, \ldots, D$ (injetoras e 3 vezes diferenciáveis) e aplicar a função g_j à j-ésima componente do produto (2.4) . Em outras palavras, supor que a função

$$\psi(x_\bullet|\beta),$$

é da forma

[2]Note que apenas o método de Hijazi-Jernigan possui um procedimento envolvendo eventos aleatórios e, portanto, apenas este produz resultados diferentes para todos os parâmetros de maximização fixados.

$$\psi(x_\bullet, \beta) = \begin{bmatrix} \lambda_1(x_\bullet) \\ \lambda_2(x_\bullet) \\ \vdots \\ \lambda_D(x_\bullet) \end{bmatrix} = \begin{bmatrix} g_1(\ m_{11}\beta_{11}x_1 + m_{12}\beta_{12}x_2 + \ldots + m_{1k}\beta_{1k}x_k\ ) \\ g_2(\ m_{21}\beta_{21}x_1 + m_{22}\beta_{22}x_2 + \ldots + m_{2k}\beta_{1k}x_k\ ) \\ \vdots \\ g_D(\ m_{D1}\beta_{D1}x_1 + m_{D2}\beta_{D2}x_1 + \ldots + m_{Dk}\beta_{Dk}x_k\ ) \end{bmatrix}, \tag{2.11}$$

$x_\bullet = (x_1, \ldots, x_k)$.

Neste modelo, a condição (2.5) pode ser eliminada. Sob as hipóteses sobre as funções $g_j, j = 1, 2, \ldots, D$ e sobre a matriz X de (1.1), os autores garantem a existência e unicidade do estimador de máxima verossimilhança, baseados em um resultado de Krätschmer (2007). Porém,o que Krätschmer mostra é que o conjunto de possíveis amostras tais que o a função de verossimilhança assume um máximo em dois ou mais pontos distintos possui medida de Lebesgue nula. Isso garante a unicidade do estimador de máxima verossimilhança com probabilidade 1, mas não garante a existência.

Naquele trabalho, os autores utilizam a função exponencial ($g_j(t) = e^t\ \forall\ j$). Em algumas de nossas simulações com os dados do Lago Ártico, porém, o uso da função exponencial apresentou muita instabilidade. Uma alternativa é utilizar a função Arcotangente que, a menos de uma constante positiva maior do que $\frac{\pi}{2}$, também satisfaz a essas condições (mais comentários no Capítulo 4, Figura 4.2).

Capítulo 3

Outros modelos de regressão

3.1 Transformações $S_D \longmapsto \mathbb{R}^{D-1}$

Segundo Aitchison (1982), uma das dificuldades iniciais para a regressão de dados composicionais é a escassez de famílias paramétricas de modelos probabilísticos definidos no $D-1$-Simplex (a mais conhecida é a Distribuição de Dirichlet). Aitchison propõe contruir novos modelos probabilísticos no Simplex utilizando distribuições conhecidas em R^d $(d = D-1)$ por meio de transformações bijetoras convenientes. As distribuições Normais-Logísticas, apresentadas em Aitchison e Shen (1980), por exemplo, são provenientes da transformação de variáveis com distribuição normal multivariada via transformação Logaritmo da razão (alr) :

$$(y_1, y_2, \ldots, y_D) \in S_D \overset{alr}{\longmapsto} (\log(y_1/y_D), \log(y_2/y_D), \ldots \log(y_{D-1}/y_D)) \in \mathbb{R}^d. \tag{3.1}$$

Essa nova classe de distribuições definidas no Simplex, porém, não é de serventia imediata para a aplicação do método descrito no Capítulo introdutório, pois "Embora existam os momentos de todas as ordens ... suas expressões integrais[1] não são redutíveis a uma forma simples"Aitchison e Shen (1980). Aitchison, sugere transformar os dados do Simplex para $\mathbb{R}^d$, aplicar o método para os dados transformados, e retornar para o Simplex. Em um dos seus estudos de caso, ele se propõe a analisar como varia a composição do solo do Lago Ártico em função da profundidade z (Aitchinson (1986)), supondo que o vetor

$$\mu(x) = \begin{bmatrix} \log(\frac{sand}{clay})(z) = \beta_{11} + \beta_{12}z + \beta_{13}z^2 + \beta_{14}\log(z) \\ \log(\frac{silt}{clay})(z) = \beta_{21} + \beta_{22}z + \beta_{23}z^2 + \beta_{24}\log(z) \end{bmatrix} = \beta \times x, \quad x = (1, z, z^2, \log(z)) \tag{3.2}$$

representa o valor esperado de uma distribuição normal bivariada $N^2(\mu(z), \Sigma)$, onde Σ independe de z. Aitchison sugere a estimação dos parâmetros β_{ij} e Σ por mínimos quadrados multivariado ou, sob a hipótese de normalidade, pela maximização da função de verossimilhança (Aitchinson (1986), pág. 160).

[1] Em particular, a expressão correspondente ao valor esperado

No caso geral, supondo que

$$alr(y_\bullet)|x_\bullet \sim N^d(\mu(x_\bullet), \Sigma), \tag{3.3}$$

a função de verossimilhança é dada por: $L(\beta ij, 1 \leq i \leq d,\ 1 \leq j \leq k, \Sigma | alr(y_{1\bullet}), \ldots alr(y_{n\bullet})) =$

$$\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right)^n \prod_{l=1}^{n} exp\left\{-(alr(y_{l\bullet}) - \mu(x_{l\bullet}))^t \times \Sigma^{-1} \times (alr(y_{l\bullet}) - \mu(x_{l\bullet}))\right\}, \tag{3.4}$$

onde $\mu(x) = \psi(x, \beta)$ possui a mesma forma de (2.2).

Vale ressaltar que:

- Para modelos com matriz de modelo completa, os estimadores (para β) de máxima verossimilhança e de mínimos quadrados, que minimiza

$$Res^2(\beta) = \sum_{i=1}^{n} || \ (alr(y_{i\bullet}) - \mu(x_{i\bullet})) \ ||^2, \tag{3.5}$$

  são idênticos e podem ser obtidos por meio da resolução de um sistema linear (ver apêndice A).

- No caso de ajuste total (onde o resíduo quadrático (3.5) é igual a zero) a função de verossimilhana (3.4), calculada nos parâmetros β_{ij} de ajuste total, depende de Σ apenas pelo inverso do seu determinante, o qual pode aproximar-se indefinidamente de zero [2]. Isso mostra que, nesse caso, a função de verossimilhança não é limitada em Σ e, portanto, a estimação de Σ por máxima verossimilhança pode, eventualmente, não fazer sentido. Determinar quais casos são patológicos, porém, parece ser uma tarefa complexa.

3.1.1 Abrangência do método

Ressaltamos que o desenvolvimento descrito acima pode ser feito para qualquer transformação de S_D para $\mathbb{R}^d$. Além da transformação alr, as seguintes transformações são apresentadas em Aitchison (1982).

[2] Um exemplo de uma sequência de matrizes positivas-definidas com essa propriedade é $\left(\frac{1}{k}I\right)_{k \in \mathbb{N}}$, onde I denota a matriz identidade.

Nome	Especificação
Multiplicativa (M_d)	$v_i = \log \frac{y_i}{1-\sum_{j=1}^{i} y_j}, i = 1, 2, \ldots, d$
Híbrida (H_d)	$v_1 = \log \frac{y_1}{1-y_1}, \quad v_i = \log \frac{y_i}{\left(1-\sum_{j=1}^{i-1} y_j\right)\left(1-\sum_{j=1}^{i} y_j\right)}, i = 2, \ldots, d$

Tabela 3.1: *Transformações logísticas elementares* $y \in S_D \longmapsto v \in \mathbb{R}^d$

Outra transformação pode ser obtida por meio da transformação *clr* (centered log-ratio):

$$y = (y_1, y_2, \ldots, y_D) \in S_D \overset{clr}{\longmapsto} (\log(y_1/g(y)), \log(y_2/g(y)), \ldots \log(y_D/g(y))) \in V, \tag{3.6}$$

apresentada em Aitchison *et al.* (2000), entre o $D-1$-Simplex e o hiperplano V, paralelo a ele, onde $g(y)$ é a média geométrica da composição y. No entanto, o contra-domínio da transformação *clr*, não é o espaço euclidiano $\mathbb{R}^d$ mas sim um subespaço de $\mathbb{R}^D$ isomorfo a ele (os vetores transformados possuem, ainda, D componentes). Para obtermos vetores em $\mathbb{R}^d$, precisamos compor a função *clr* com algum isomorfismo linear T entre V e R^d como, por exemplo, a função que retorna as coordenadas de um elemento de V com relação à uma certa base B. Uma possível escolha para B é:

$$B = \left\{ \frac{v_1}{||v_1||}, \frac{v_2}{||v_2||}, \ldots, \frac{v_d}{||v_d||} \right\}, \; v_j = (1, \ldots, 1, \underbrace{-j}_{\text{posição } j+1}, 0, \ldots, 0) \tag{3.7}$$

Porém, simulações com os dados do Lago Ártico geraram as mesmas funções de previsão, trocando-se a transformação *alr* pela transformação $T \circ clr$ (para o isomorfismo T relativo à base (3.7)). Determinar as condições sob as quais transformações distintas levam a resultados idênticos caracteriza, portanto, uma questão de fundamental importância para determinar a abrangência do método descrito na seção anterior. Aitchison dá o primeiro avanço nessa direção, ao mostrar que a função de previsão (para a transformação *alr*) é invariante sobre permutações das componentes das composições (ver Aitchinson (1986), pág. 161). Nós obtivemos uma resposta geral para essa questão:

Definição: Duas transformações $T_1, T_2 : S_D \longrightarrow \mathbb{R}^d$ são equivalentes se $T_1(y_\bullet)|x_\bullet \sim N(\mu_1(x_\bullet) = \beta_1 \times x_\bullet \;, \; \Sigma_1)$ sempre que $T_2(y_\bullet)|x_\bullet \sim N(\mu_2(x_\bullet) = \beta_2 \times x_\bullet \;\; \Sigma_2)$ e vice-versa.

Teorema 1. *Duas transformações* $T_1, T_2 : S_D \longrightarrow R^d$ *são equivalentes se e somente se* $T_1 \circ T_2^{-1} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ *é uma transformação linear. Nessas condições, as funções de previsão* $Prev_{T_1}(x_\bullet) = T_1^{-1}(\widehat{\beta}_1 \times x_\bullet)$ *and* $Prev_{T_2}(x_\bullet) = T_2^{-1}(\widehat{\beta}_2 \times x_\bullet)$*, dadas pelas estimativas de máxima verossimilhança* $\widehat{\beta}_1$ *e* $\widehat{\beta}_2$*, são idênticas (para modelos com matriz de modelo completa).*

Prova:

Suponha que $T_1 \circ T_2^{-1} : U \longmapsto M \times U$ é linear (M denota a matriz da transformação $T_1 \circ T_2^{-1}$ com relação à base canônica de $\mathbb{R}^d$).

É um fato conhecido sobre a distribuição Normal, que

$$T_2(y_\bullet)|x_\bullet \sim N(\beta_2 \times x_\bullet, \Sigma_2) \Longleftrightarrow \underbrace{M \times T_2(x_\bullet)}_{T_1(x_\bullet)}|x_\bullet \sim N(M \times \beta_2 \times x_\bullet \;\; M \times \Sigma_2 \times M^t).$$

Por (A.5), temos que

$$\sum_{l=1}^{n} x_{lj}(T_1(y_{l\bullet}) - \widehat{\beta_1} \times x_{l\bullet} \;=\; 0, \; j = 1, 2, \ldots, k. \tag{3.8}$$

$$\sum_{l=1}^{n} x_{lj}(T_2(y_{l\bullet}) - \widehat{\beta_2} \times x_{l\bullet} \;=\; 0, \; j = 1, 2, \ldots, k. \tag{3.9}$$

Substituindo $T_1(y_{l\bullet}) \;=\; T_1 \circ T_2^{-1}(T_2(y_{l\bullet})) \;= M \times T_2(y_{l\bullet})$ em (3.8), obtemos

$$\sum_{l=1}^{n} x_{lj}(T_2(y_{l\bullet}) - M^{-1} \times \widehat{\beta_1} \times x_{l\bullet} \;=\; 0, \; j = 1, 2, \ldots, k. \tag{3.10}$$

Comparando (3.9) e (3.10), devemos[3] ter $\widehat{\beta_1} = M \times \widehat{\beta_2}$ e, então,

$$Prev_{[T_1]}(x_\bullet) = \; T_1^{-1}(M \times \widehat{\beta_2} \times x_\bullet) \;=\; T_2^{-1}(\widehat{\beta_2} \times x_\bullet) \;=\; Prev_{[T_2]}(x_\bullet).$$

Por outro lado, suponha que T_1 e T_2 são equivalentes e seja $\Phi = T_1 \circ T_2^{-1}$.

Tome $k = d$ e $f_1, f_2, \ldots, f_d$, definidas no capítulo introdutório, tais que $x_\bullet = z_\bullet \; \forall \; z_\bullet \; \in \; \Omega$.

Por hipótese, para $w(x_\bullet) = T_1(y_\bullet)|x_\bullet \sim N(x_\bullet, \Sigma_1)$ ($\beta_1 =$ matriz identidade), devemos ter

$$\zeta = \zeta(x_\bullet) = T_2(y_\bullet)|x_\bullet \sim N(\beta_2 \times x_\bullet, \Sigma_2), \tag{3.11}$$

para certas matrizes $\beta_2, \; \Sigma_2$ ($\forall \; x_\bullet \; \in \; \Omega$).

[3]Pois $M \times \widehat{\beta_2}$ também satisfaz (A.5).

Além disso, se $w(x_\bullet)$ possui *f.d.p.* (função densidade de probabilidade) $f(w|x_\bullet)$, então $\zeta(x_\bullet)$ possui *f.d.p.* $h(\zeta|x_\bullet) = f(\Phi(\zeta)|x_\bullet)\left|\frac{\partial\Phi}{\partial\zeta}\right|$.

Por (3.11), $h(\zeta|x_\bullet)$ é simétrica com relação a ζ e $\beta_2 \times x_\bullet$. Logo, seu gradiente (como função de $x_\bullet$) deve se anular em $x_\bullet^* = \beta_2^{-1} \times \zeta$. Portanto,

$$0 \;=\; \nabla h(x_\bullet^*|\zeta) \times \beta_2 \;=\; \nabla f(x_\bullet^*|\Phi(\zeta)) \times \underbrace{\left|\frac{\partial\Phi}{\partial\zeta}(\zeta)\right|}_{\neq 0}. \tag{3.12}$$

Segue que $x_\bullet^* = \Phi(\zeta)$, isto é: $\Phi(\zeta) \;=\; \beta_2^{-1} \times \zeta \;(\;\forall\; \zeta \;\in\; \mathbb{R}^d)$

□

Corolário 1. *1. alr e alr $\circ$ P são equivalentes para qualquer permutação P das componentes das composições.*

2. $T \circ clr$ e alr são equivalentes para todo isomorfismo linear $T : V \longrightarrow \mathbb{R}^d$

3. As Transformações logísticas Multiplicativa (M_d) e Híbrida (H_d) são equivalentes somente para $D \leq 3$.

Prova:

1. Sejam $T_1 = alr \circ P$ e $T_2 = alr$. Sem perda de generalidade, vamos supor que P é uma transposição, isto é: $P(y_1, \ldots, y_i, \ldots, y_j, \ldots, y_D) = (y_1, \ldots, y_j, \ldots, y_i, \ldots, y_D)$.

   Para $u = T_2(y) \;\in\; R^d$, temos:

   1º caso: $i < j < D$.

$$\begin{aligned} T_1(T_2^{-1}(u)) &= T_1(y) \\ &= alr((y_1, \ldots, y_j, \ldots, y_i, \ldots, y_D)) \\ &= (\log(y_1/y_D), \ldots \log(y_j/y_D), \ldots, \log(y_i/y_D), \ldots \log(y_{D-1}/y_D)) \\ &= (u_1, \ldots, u_j, \ldots, u_i, \ldots, u_d) \end{aligned}$$

   2º caso: $i < j = D$.

$$\begin{aligned} T_1(T_2^{-1}(u)) &= T_1(y) \\ &= alr((y_1, \ldots, y_j, \ldots, y_{D-1}, y_i)) \\ &= (\log(y_1/y_i), \ldots, \log(y_j/y_i), \ldots, \log(y_{D-1}/y_i)) \\ &= (u_1 - u_i, \ldots, u_j - u_i, \ldots, u_d - u_i) \end{aligned}$$

2. Para $u = alr(Y)$, temos

$$T(clr(Y)) = T\left(\frac{D-1}{D}u_1 - \frac{1}{D}\sum_{j\neq 1} u_j,\ \ldots,\ \frac{D-1}{D}u_d - \frac{1}{D}\sum_{j\neq d} u_j, -\frac{1}{D}\sum_{j=1}^{d} u_j\right).$$

3.
   - Para $D = 2$, $M_d(Y) = H_d(Y)\ \forall\ Y\ \in\ S_D$.
   - Para $D = 3$ e $M_d(Y) = u$, temos $H_d(Y) = (u_1, u_2 - u_1)$.
   - Para $D > 3$, $M_d(Y) = u$ e $H_d(Y) = w$, temos que $w_3 = u_3 + u_2 - \log(y_2)$, e $\log(y_2)$ não pode ser expressa como combinação linear de $u_1, u_2, \ldots, u_d$.

□

3.2 Coordenadas esféricas

Wang *et al.* (2007) sugeriram um modelo de regressão baseado em outra parametrização do Simplex por coordenadas esféricas[4]:

$$(\alpha_1, \alpha_2, \ldots, \alpha_d) \ \in\ \left(]0, \tfrac{\pi}{2}]\right)^d \ \longmapsto y\ \in \overline{S_D}$$

$$\begin{aligned} \gamma_1 &= \prod_{j=1}^{d} sin(\alpha_j), \\ \gamma_i &= cos(\alpha_{i-1}) \prod_{j=i}^{d} sen(\alpha_j), \quad i = 2, 3, \ldots, D \\ y_j &= \gamma_j^2, \quad j = 1, 2, \ldots, D. \end{aligned} \tag{3.13}$$

Nesse modelo, os dados transformados, correspondentes às coordenadas angulares $\alpha_1, \alpha_2, \ldots, \alpha_d$, são modelados diretamente em função das covariáveis por algum método de inferência. Segundo os autores, esse modelo apresenta uma vantagem sobre o método de Aitchison, já que, como não há divisões, o modelo comporta dados composicionais com componentes eventualmente nulas. Infelizmente, porém, o problema de lidar com componentes nulas (o qual nos abstemos nesta apresentação) não nos parece ter uma solução tão simples.

Segundo os autores (que se limitam ao uso de covariáveis unidimensionais, mais precisamente o tempo), o modelo consiste em construir d modelos, um para cada coordenada angular:

$$\theta_j(t) \ = \ h_j(t) + \epsilon_j(t), \quad j = 1, 2, \ldots, d, \tag{3.14}$$

onde $h_j(t), j = 1, 2, \ldots, d$ são as funções que descrevem os modelos e $\epsilon_j(t)$ são os erros observados.

[4] O qual suporta componentes nulas nas composições.

Naquele trabalho, os autores utilizam funções polinomiais para as funções $h_j, j = 1, 2, \ldots, d$. Mais em geral, podemos supor que tais funções são combinações lineares de funções de livre escolha, como assumido nos Capítulos anteriores, isto é:

$$\begin{bmatrix} h_1(t) \\ h_2(t) \\ \vdots \\ h_d(t) \end{bmatrix} = \begin{bmatrix} m_{11} * \beta_{11} & m_{12} * \beta_{12} & \ldots & m_{1k} * \beta_{1k} \\ m_{21} * \beta_{21} & m_{22} * \beta_{22} & \ldots & m_{2k} * \beta_{1k} \\ \vdots & \vdots & & \vdots \\ m_{D1} * \beta_{d1} & m_{d2} * \beta_{d2} & \ldots & m_{dk} * \beta_{dk} \end{bmatrix} \times \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_k(t) \end{bmatrix}, \tag{3.15}$$

onde $f_1(t), f_2(t), \ldots, f_k(t)$ e $M_{mod} = (m_{ij})$ são definidas nas Seções 1.1 e 2.2, respectivamente. Sob tais hipóteses, podemos listar alguns inconvenientes na implementação e aplicação desse modelo de regressão:

1. Uma primeira dificuldade diz respeito ao método de inferência utilizado para a estimação dos coeficientes β_s. A escolha mais natural, certamente, é a estimação por mínimos quadrados. Porém, assim como no modelo de Dirichlet linear, o processo de otimização deve ser condicionado, já que os ângulos devem sempre pertencer ao intervalo $[0, \pi/2]$. Por outro lado, a função a ser minimizada (a soma dos quadrados dos resíduos) é uma forma quadrática positiva definida e, portanto, assume um único ponto de mínimo (o único que anula o gradiente). Consequentemente, se um tal ponto de mínimo não se encontra na região viável, qualquer algoritmo de maximização condicionada resultará em um ponto da fronteira. Em outras palavras, existirão índices i e j para os quais teremos $h_j(t_i) = 0$ ou $h_j(t_i) = \pi/2$ e, portanto, há grande chances de que $h_j(t) < 0$ ou $h_j(t) > \pi/2$, para t próximo à t_i, o que tornaria o modelo inutilizável.

2. Ironicamente, esse inconveniente se torna ainda mais provável nos casos para os quais o modelo foi desenvolvido, isto é, nas aplicações onde o conjunto de dados possui composições com algumas componentes nulas. Pois, para qualquer método de inferência cujo objetivo seja a aproximação das funções $h_1, h_2, \ldots, h_d$ aos dados observados, componentes nulas gerarão funções assumindo valores próximos à região da fronteira.

Não fomos capazes de reproduzir os resultados apresentados pelos autores na análise de seus conjuntos de dados (vale ressaltar que o método de inferência utilizado não foi explicitado).

O primeiro conjunto de dados, constituído por doze composicões com quatro componentes cada, foi gerado para simulações (a covariável assumindo os valores $1, 2, \ldots, 12$). O segundo, constituído por doze composicões com três componentes cada, é referente à variação das componentes da economia chinesa (setor primário, secundário e terciário) entre os anos de 1990 e 2001. Considerando modelos polinomiais, os autores expõem as seguintes estimativas

$$\begin{array}{l} \theta_1^*(t) = 1.4289 + 0.0304t^2 - 0.0009t^3 \\ \theta_2^*(t) = 1.6346 - 0.1263t \\ \theta_3^*(t) = 1.4546 + 0.1149t - 0.0317t^2 + 0.0013t^3 \end{array} \quad \text{e} \quad \begin{array}{l} \theta_1'(t) = 1.0572 - 0.0161t + 0.0007t^2 \\ \theta_2'(t) = 1.1572 - 0.0214t + 0.0008t^2 \end{array}. \tag{3.16}$$

Nós obtivemos as seguintes estimativas de mínimos quadrados (por maximização irrestrita):

$$\begin{array}{l}
\theta_1^a(t) = (9.864e-1) - (3.627e-2)t^2 + (2.753e-3)t^3 \\
\theta_2^a(t) = (1.635e+0) - (1.264e-1)t \\
\theta_3^a(t) = (1.486e+0) + (1.012e-1)t - (2.983e-2)t^2 + (1.221e-3)t^3 \\
\\
\theta_1^b(t) = (2.873e+3) - (2.871e+0)t + (7.178e-4)t^2 \\
\theta_2^b(t) = (3.077e+3) - (3.071e+0)t + (7.666e-4)t^2
\end{array} . \tag{3.17}$$

Para o primeiro conjuntos de dados, a estimativa de mínimos quadrados não encontra-se dentro da região viável ($\theta_3^a(2) = 1.579 > \pi/2$).

3.3 Outras transformações no Simplex

Motivados pelo método geral proposto na Seção 3.1.1, fomos à procura de mais transformações entre o Simplex e o Espaço euclidiano. De forma independente[5], encontramos uma série de outras transformações baseadas em coordenadas esféricas generalizadas.

Ao contrário das transformações da tabela 3.1, a transformação *alr* possui uma fácil interpretação geométrica, como mostrado na figura 3.1.

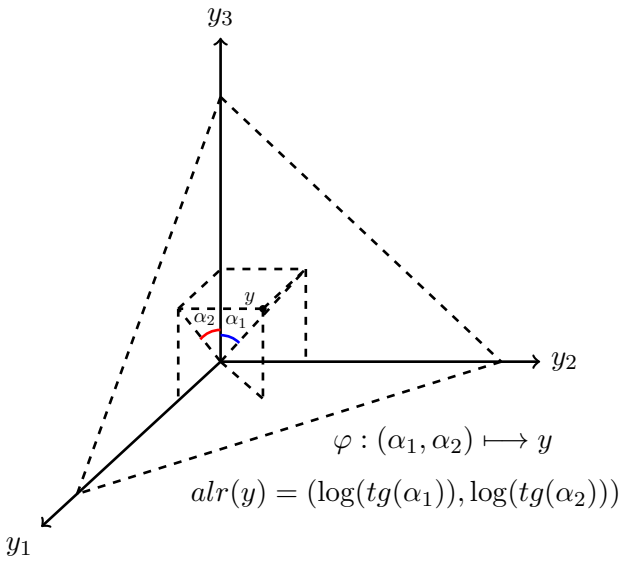


Figura 3.1: *Interpretação geométrica da transformação alr*

[5] O trabalho de Wang só foi conhecido posteriormente.

Como podemos observar, essa transformação é baseada na parametrização φ, do Simplex, em termos das coordenadas angulares periféricas α_1 e α_2. Mais transformações são obtidas ao considerarmos a parametrização do Simplex por coordenadas angulares centrais (coordenadas esféricas), como exemplificado na figura 3.2, para $D = 3$.

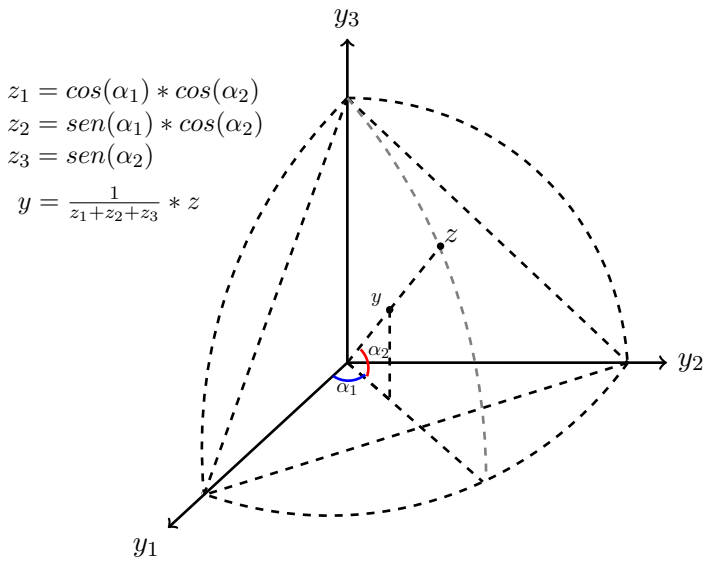


Figura 3.2: *Parametrização do 3-Simplex por coordenadas esféricas*

Essa parametrização pode ser extendida para $D > 3$, utilizando coordenadas esféricas generalizadas:

$$\varphi : \ (\alpha_1, \alpha_2, \ldots, \alpha_d) \ \in \ \left(]0, \tfrac{\pi}{2}[\right)^d \ \longmapsto y \ \in S_D$$

$$\begin{aligned} z_1 &= \prod_{j=1}^{d} cos(\alpha_j), \\ z_i &= sen(\alpha_{i-1}) \prod_{j=i}^{d} cos(\alpha_j), \quad i = 2, 3, \ldots, D \\ y &= z \Big/ \sum_{j=1}^{d} z_j \end{aligned} \tag{3.18}$$

e a sua inversa

$$z \ = \ y/||y|| \quad \alpha_i \ = \ tg^{-1}\left(\frac{z_{i+1}}{\sqrt{\sum_{j=1}^{i} z_j^2}} \right) \tag{3.19}$$

Obtendo-se as cordenadas angulares $\alpha_1, \alpha_2, \ldots, \alpha_d$, podemos compor novas transformações entre o Simplex e o Espaço Euclidiano, aplicando-se as funções tangente e/ou logaritmo (ou outras funções adequadas). Note que há relação biunívoca entre os modelos baseados em coordenadas angulares periféricas e centrais. A transformação correspondente à Logaritmo da razão, em coordenadas esféricas, é a transformação Log-esférica , dada por:

$$\begin{aligned} y \in S_D &\overset{logEsf}{\longmapsto} (\log(tg(\alpha_1)), \log(tg(\alpha_2)), \ldots, \log(tg(\alpha_d))) \in \mathbb{R}^d, \\ z = y/||y||, \quad \alpha_i &= tg^{-1}\left(\frac{z_{i+1}}{\sqrt{\sum_{j=1}^{i} z_j^2}}\right). \end{aligned} \tag{3.20}$$

Aplicando-se a função tangente diretamente, obtemos uma nova classe de transformações:

Definição: A transformação Tangente-esférica com parâmetros $h_1, h_2, \ldots, h_d :]0, \frac{\pi}{2}[\longrightarrow]-\frac{\pi}{2}, \frac{\pi}{2}[$ é definida por

$$y \in S_D \overset{TgEsf}{\longmapsto} (tg(h_1(\alpha_1)), tg(h_2(\alpha_2)), \ldots, tg(h_d(\alpha_d))) \in \mathbb{R}^d, \tag{3.21}$$

onde $\alpha_1, \alpha_2, \ldots, \alpha_d$ são dados em (3.20).

A transformação correspondente à tangente-esférica com parâmetros lineares ($h_j(x) = 2x - \pi/2, \ \forall\ j$), em coordenadas angulares periféricas, é:

$$y \in S_D \overset{TgRatio}{\longmapsto} z \in \mathbb{R}^d, \quad z_j = tg\left(2 * arctg\left(\frac{y_j}{y_D}\right) - \frac{pi}{2}\right), j = 1, 2, \ldots, d \tag{3.22}$$

Utilizando a relação $tg(2x - \pi/2) = \frac{1}{2}\left(tg(x) - \frac{1}{tg(x)}\right)$, (3.22) assume a forma simplificada

$$TgRatio(y) = \frac{1}{2}\left(\frac{y_1}{y_D} - \frac{y_D}{y_1}, \frac{y_2}{y_D} - \frac{y_D}{y_2}, \ldots, \frac{y_d}{y_D} - \frac{y_D}{y_d}\right) \tag{3.23}$$

Proposição 1. *$T_{logEsf} \circ P_1$ e $T_{logEsf} \circ P_2$ são equivalentes para duas permutações P_1 e P_2 se e somente se $P_1\ P_2^{-1}$ permuta apenas as duas primeiras componentes das composições (análogo para T_{tgEsf} com parâmetros lineares).*

Prova:

Para $z = \frac{y}{||y||}$, temos que $T_{logEsf}(y) = \left(\log\frac{z_2}{z_1}, \log\frac{z_3}{\sqrt{\sum_{j=1}^{2} z_j^2}}, \ldots, \log\frac{z_D}{\sqrt{\sum_{j=1}^{d} z_j^2}}\right)$ e, claramente, uma permutação que represente uma transformação linear deve fixar todas as somas sob as raízes

quadradas.

□

Corolário 2. *$\{T_{logEsf} \circ P : P$ é uma permutação de $1, 2, \ldots, D\}$ possui $\frac{D!}{2}$ classes de equivalência distintas (o mesmo para T_{tgEsf} com parâmetros lineares).*

3.3.1 Observações

Estatísticos frequentemente procuram procedimentos invariantes sobre mudanças de coordenadas. Sendo assim, alguém poderia questionar o uso das transformações esféricas, já que elas produzem resultados diferentes permutando-se as componentes das composições. Porém, devemos ressaltar que essa mudança de coordenadas não representa apenas um artifício para simplificação de cálculos, mas sim uma mudança sobre todas as hipóteses estatísticas consideradas (de acordo com a definição de transformações equivalentes). É claro que, em cada aplicação, quanto mais estreita a relação entre as hipóteses estatísticas e a natureza dos dados, maior a adequação do modelo. Portanto, se os dados seguem o modelo linear Logaritmo da razão (em um sentido estrito), não será possível obter resultados melhores pela aplicação de outros modelos. Porém, se assumimos que cada hipótese estatística possui a mesma probalidade de ocorrer (assumirmos uma priori uniforme), então provavelmente encontraremos um modelo melhor entre os $\frac{D!}{2}$ gerados por permutações dos modelos esféricos do que o único modelo gerado pelas permutações aplicadas à transformação Logaritmo da razão. Esse parece ser o caso para as composições do Lago Ártico, como mostrado no Capítulo 4, figura 4.3. O melhor modelo Log-esférico apresentou um ajuste mais adequado do que o ajuste pela transformação *alr* e os outros modelos Log-esféricos tampouco apresentaram pior desempenho.

O modelo Tangente-esférico (com parâmetros lineares) não gerou bons resultados para os mesmos parâmetros fixados nessa análise. Porém, com um pouco mais de insistência, conseguimos produzir (artificialmente) funções f_js de modo a gerar resultados adequados, como mostrado na Figura 4.4.

3.4 Método dos mínimos quadrados no Simplex

Por meio da transformação *clr* (3.6) e da estrutura euclidiana de $\mathbb{R}^D$, Aitchison introduziu no $D-1$-Simplex uma estrutura de $\mathbb{R}$-espaço vetorial[6] e um produto interno do qual deriva a distância

$$\Delta(y, w) = \left[\sum_{j=1}^{D} \left(\log \frac{y_j}{g(y)} - \log \frac{w_j}{g(w)} \right)^2 \right]^{1/2}, \; g = \text{média geométrica} \tag{3.24}$$

[6] Mais em geral, se $(V, +, .)$ é um espaço vetorial e $\varphi : V \longmapsto B$ é uma função bijetora, então φ define um isomorfismo de espaços vetoriais entre V e B, com as operações em $B \oplus : B \times B \longmapsto B$, $\bullet : \mathbb{R} \times B \longmapsto B$ definidas por $b_1 \oplus b_2 = \varphi(\, \varphi^{-1}(b_1) + \varphi^{-1}(b_2)\,)$, $\alpha \bullet b = \varphi(\, \alpha.\varphi^{-1}(b)\,)$.

(conhecida como a Distância de Aitchison) que "possui todas as propriedades necessárias para aplicações em análise de dados composicionais ", Aitchison *et al.* (2000). Da mesma forma,

$$\left(\sum_{i=1}^{n} \Delta(Y_i, Z_i)^2\right)^{1/2} \tag{3.25}$$

define uma distância em $(S_D)^n$ que, segundo os critérios de Aitchison, poderia ser considerada como favorita entre as demais distâncias. Logo, observados $(x_{1\bullet}, y_{1\bullet}), (x_{2\bullet}, y_{2\bullet}), \ldots (x_{n\bullet}, y_{n\bullet})$ e dada uma família paramétrica de modelos de regressão $\{h(|\beta) : \beta \in B\}$, podemos pensar em um método de ajuste do tipo mínimos quadrados, isto é: escolher o modelo que minimize a quantidade

$$R^2(\beta) = \sum_{i=1}^{n} \Delta(y_{i\bullet}, h(x_{i\bullet}|\beta))^2. \tag{3.26}$$

Para variáveis assumindo valores no $D-1$-Simplex, modelos da forma $g = \left(\frac{g_1}{\sum g_j}, \frac{g_2}{\sum g_j}, \ldots, \frac{g_D}{\sum g_j}\right)$, onde $g_1(x_\bullet), g_2(x_\bullet), \ldots, g_D(x_\bullet)$ são funções positivas da covariável $x_\bullet$, representam candidatos naturais. De fato, se $y_1, y_2, \ldots, y_D$, são variáveis positivas de mesma espécie modeladas, respectivamente, por $g_1(x_\bullet), g_2(x_\bullet), \ldots, g_D(x_\bullet)$, então a composição das frequências relativas $y = \left(\frac{y_1}{\sum y_j}, \frac{y_2}{\sum y_j}, \ldots, \frac{y_D}{\sum y_j}\right)$ é modelada por por g.

Capítulo 4

Aplicações

Os modelos de regressão descritos nos Capítulos anteriores foram implementados utilizando o ambiente de programação *R* 2.12.2. Para maximização restrita às condições (2.5) foi utilizada a função *constrOptim*() e, para otimização/minimização irrestrita, a função *optim*(), exceto para os ajustes de mínimos quadrados linear, para os quais desenvolvemos uma rotina baseada no Teorema do Apêndice A.

Testamos a performance dos modelos propostos para os três conjuntos de dados descritos no Capítulo introdutório, variando os parâmetros (a forma da variável estendida, entre outros). No geral, buscamos os parâmetros que produzissem os melhores ajustes (avaliados visualmente por meio de gráficos) para ilustrar a flexibilidade dos modelos. Para alguns casos, foram necessárias mais do que funções polinomiais (a escolha mais simples e usual).

A seguir apresentamos os parâmtros ótimos (de máxima verossimilhança e de mínimos quadrados) para cada caso, juntamente com o gráfico do ajuste correspondente.

Ao final deste Capítulo encontra-se uma discussão e alguns questionamentos sobre a definição do espaço paramétrico para alguns dos modelos descritos.

4.1 Lago Ártico

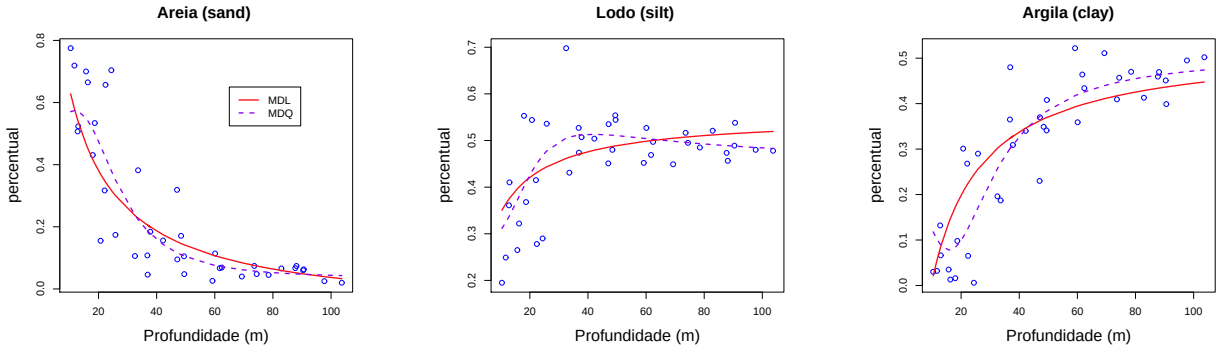
4.1.1 Modelo de Dirichlet Linear

A Figura 4.1 mostra a previsão gerada pelos Modelos de Dirichlet Lineares de primeira (MDL) e segunda (MDQ) ordem[1]:

$$\begin{array}{ll} \lambda_1(z_\bullet) = \beta_{11} + \beta_{12} z_\bullet & \lambda_1(z_\bullet) = \beta_{11} + \beta_{12} z_\bullet + \beta_{13} z_\bullet^2 \\ \lambda_2(z_\bullet) = \beta_{21} + \beta_{22} z_\bullet & \lambda_2(z_\bullet) = \beta_{21} + \beta_{22} z_\bullet + \beta_{23} z_\bullet^2 \\ \lambda_3(z_\bullet) = \beta_{31} + \beta_{32} z_\bullet & \lambda_3(z_\bullet) = \beta_{31} + \beta_{32} z_\bullet + \beta_{33} z_\bullet^2 \\ \quad MDL & \quad MDQ \end{array} \qquad (z_\bullet = \text{profundidade})$$

correspondentes aos parâmetros de máxima verossimilhança da Tabela 4.1. Para esses modelos, temos, respectivamente: MDL: $k = 2$, $x_\bullet = (1, z_\bullet)$, MDQ: $k = 3$, $x_\bullet = (1, z_\bullet, z_\bullet^2)$.

[1] Estes modelos também são abordados em Hijazi e Jernigan (2009).



(a) (b) (c)

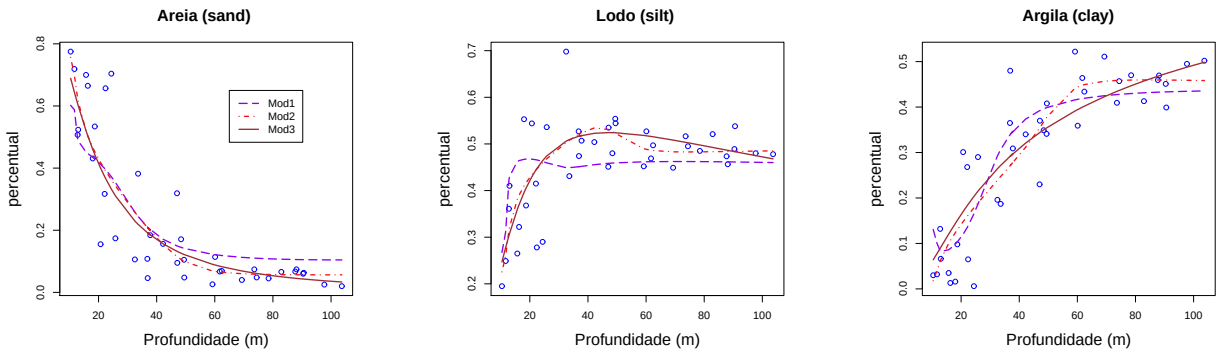
Figura 4.1: *Modelos de Dirichlet lineares*

		MDL			MDQ		
$\widehat{\beta} =$	β_{11} β_{12} β_{13}	$4.851e+0$	$-3.262e-2$	$-$	$5.226e+0$	$-7.134e-2$	$1.051e-3$
	β_{21} β_{22} β_{23}	$2.237e-1$	$2.206e-1$	$-$	$3.466e+0$	$-2.085e-1$	$1.121e-2$
	β_{31} β_{32} β_{33}	$-2.049e+0$	$2.117e-1$	$-$	$3.692e+0$	$-3.971e-1$	$1.283e-2$

Tabela 4.1: *Parâmetros de máxima verossimilhança*

4.1.2 Modelo de Melo, Vasconcellos e Lemonte

A Figura 4.2 mostra a previsão gerada por três instâncias distintas do Modelo de Dirichlet de Melo, Vasconcellos e Lemonte. A primeira (Mod1) corresponde ao uso da função positiva $1.6 + tg^{-1}(u)$. Na segunda (Mod2), aplicamos uma dilatação a esta função para melhorar o ajuste. Na terceira (Mod3), utilizamos a função Exponencial, modificando a forma de $x_\bullet$, naturalmente. É importante mencionar que, no caso da função Exponencial, o processo iterativo de otimização não foi completado, pois o valor da função de verossimilhança correspondente excedeu a precisão numérica do R (10^{307}).



(a) (b) (c)

Figura 4.2: *Modelos de Melo, Vasconcellos e Lemonte.*

Legenda:

$z_\bullet$ = profundidade.

Mod1: $k = 3,\ x_\bullet = (1, \log(0.1 * z_\bullet), \log(0.1 * z_\bullet)^{0.5}$
$g_j(u) = 1.6 + tg^{-1}(u),\ j = 1, 2, 3.$

Mod2: $k = 3,\ x_\bullet = (1, \log(0.1 * z_\bullet), \log(0.1 * z_\bullet)^{0.5}$
$g_j(u) = 10 * [1.6 + tg^{-1}(u)],\ j = 1, 2, 3.$

Mod3: $k = 2,\ x_\bullet = (1, \log(z_\bullet))$
$g_j(u) = exp(u),\ j = 1, 2, 3.$

	$\widehat{\beta} =$	$\begin{matrix}\beta_{11} & \beta_{12} & \beta_{13}\\ \beta_{21} & \beta_{22} & \beta_{23}\\ \beta_{31} & \beta_{32} & \beta_{33}\end{matrix}$	
Mod1	$1.633e+1$	$7.612e+0$	$-2.311e+1$
	$2.991e+0$	$4.432e+1$	$-2.506e+1$
	$-2.850e-1$	$6.786e+0$	$-6.668e+0$
Mod2	$2.705e+0$	$3.721e+0$	$-9.541e+0$
	$5.196e-1$	$8.007e+0$	$-1.050e+1$
	$-4.829e+1$	$-1.871e+1$	$6.151e+1$
Mod3	$2.482e-1$	$3.300e-1$	$-$
	$-4.512e+0$	$1.924e+0$	$-$
	$-7.299e+0$	$2.538e+0$	$-$

Tabela 4.2: *Parâmetros de máxima verossimilhança*

4.1.3 Modelo linear Logaritmo da Razão/ Log-esférico

A Figura 4.3 compara as performances dos Modelos Logaritmo da Razão (Mod1), do Modelo Log-esférico (Mod3) e dos modelos obtidos pela aplicação do Modelo Log-esférico após permutar as componentes das composições. O ajuste gerado pelo Modelo Mod2 foi ligeiramente superior ao ajuste gerado pela transformação *alr* (Mod1) e os outros modelos tampoco apresentaram ajustes de qualidade inferior.

Legenda:

$z_\bullet$ = profundidade, $k = 4,\ x_\bullet = (1,\ z_\bullet,\ z_\bullet^2,\ \log(z_\bullet))$

Mod1 : $T = alr$ Mod2 : $T = T_{logEsf} \circ P,\quad P(y_1, y_2, y_3) = (y_1, y_3, y_2)$

Mod3 : $T = T_{logEsf}$ Mod4 : $T = T_{logEsf} \circ P,\quad P(y_1, y_2, y_3) = (y_3, y_2, y_1)$

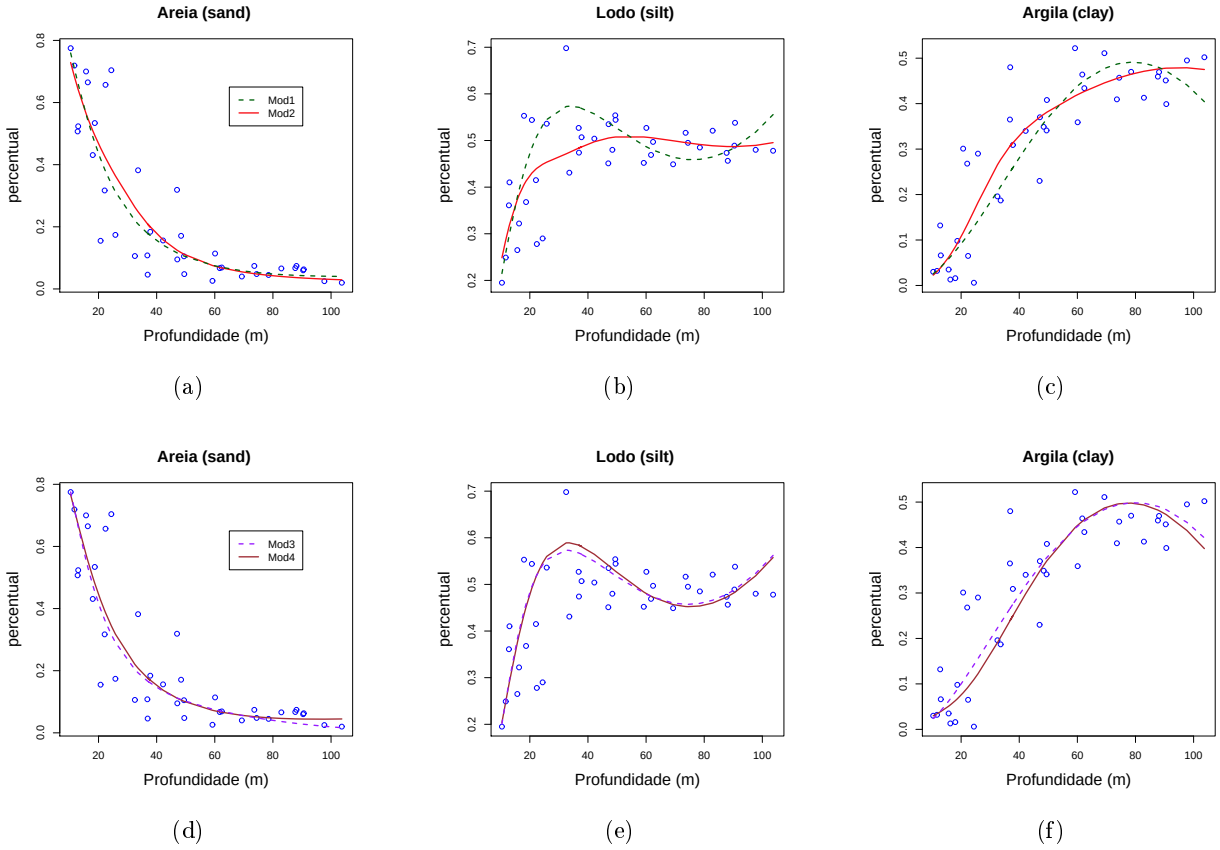


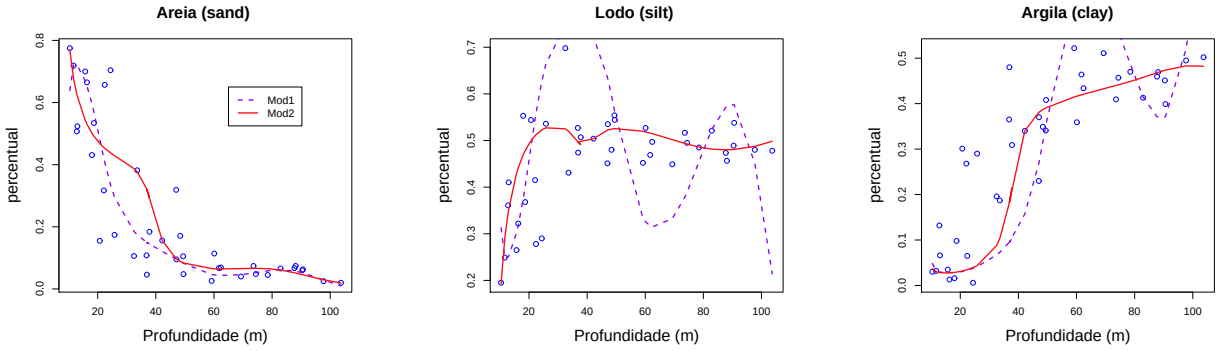
Figura 4.3: *Modelo linear Log-esférico vs Modelo linear Logaritmo da razão*

	$\widehat{\beta} =$	$\begin{matrix}\beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24}\end{matrix}$		
Mod1	$8.794e+0$	$-6.619e-2$	$4.851e-4$	$-2.037e+0$
	$2.060e+0$	$-9.804e-2$	$5.954e-4$	$4.356e-1$
Mod2	$-1.054e+1$	$9.998e-3$	$-1.439e-4$	$2.979e+0$
	$-6.441e+0$	$-9.517e-2$	$3.562e-4$	$2.697e+0$
Mod3	$-6.956e+0$	$-2.885e-2$	$6.807e-5$	$2.529e+0$
	$-6.193e+0$	$6.684e-2$	$-5.190e-4$	$9.698e-1$
Mod4	$2.556e+0$	$-8.548e-2$	$5.394e-4$	$1.715e-1$
	$7.988e+0$	$6.635e-2$	$-3.874e-4$	$-3.114e+0$

Tabela 4.3: *Parâmetros de máxima verossimilhança*

4.1.4 Modelo linear Tangente-esférico

A Figura 4.4 mostra as estimativas geradas pelo Modelo Tangente-esférico correspondente às estimativas de máxima verossimilhança da Tabela 4.4. Nota-se claramente que a forma de $x_\bullet$, assim como no Modelo de Dirichlet de Melo Vasconcellos e Lemonte, foi escolhida artificialmente (modelos polinomiais não geraram bons resultados). Ainda assim, a propriedade de sensibilidade à permutações das composições (ver Corolário 2) foi fundamental para conseguirmos um bom ajuste, pois o Modelo Tangente-esférico (Mod1) não foi suficiente.



(a) (b) (c)

Figura 4.4: *Modelos Tangente-esféricos*

Legenda:

$z_\bullet$ = profundidade.

$$k = 5, \; x_\bullet = (1, \; z_\bullet, \; z_\bullet^2, (-z_\bullet + 20)^5, \quad log(z_\bullet - 10))$$

$$\text{Mod1}: \quad T = T_{TgEsf}$$
$$\text{Mod2}: \quad T = T_{TgEsf} \circ P, \quad P(y_1, y_2, y_3) = (y_3, y_1, y_2)$$

	$\widehat{\beta} =$	$\begin{matrix}\beta_{11} \\ \beta_{21}\end{matrix}$	$\begin{matrix}\beta_{12} \\ \beta_{22}\end{matrix}$	$\begin{matrix}\beta_{13} \\ \beta_{23}\end{matrix}$	$\begin{matrix}\beta_{14} \\ \beta_{24}\end{matrix}$	$\begin{matrix}\beta_{15} \\ \beta_{25}\end{matrix}$
Mod1	$-4.978e+0$	$3.850e-1$	$-3.486e-3$	$-3.980e-9$	$-6.277e-1$
	$-2.290e+1$	$1.241e+0$	$-9.813e-3$	$-4.870e-9$	$-4.110e+0$
Mod2	$2.066e+1$	$-9.619e-1$	$8.216e-3$	$6.718e-9$	$1.291e+0$
	$-1.019e+0$	$-2.368e-2$	$1.676e-5$	$-8.057e-11$	$6.598e-1$

Tabela 4.4: *Parâmetros de máxima verossimilhança*

4.2 Despesas domésticas (bivariado)

Analisamos a variação dessas composiçoes em função das covariáveis sexo (s = 0, para masculino e s = 1, para feminino) e renda (r), isto é: $z_\bullet = (s, r)$.

4.2.1 Modelo de Dirichlet Linear

A Figura 4.5 mostra o ajuste obtido pelo Modelo de Dirichlet Linear corresponde aos parâmetros da máxima verossimilhança da Tabela 4.5.

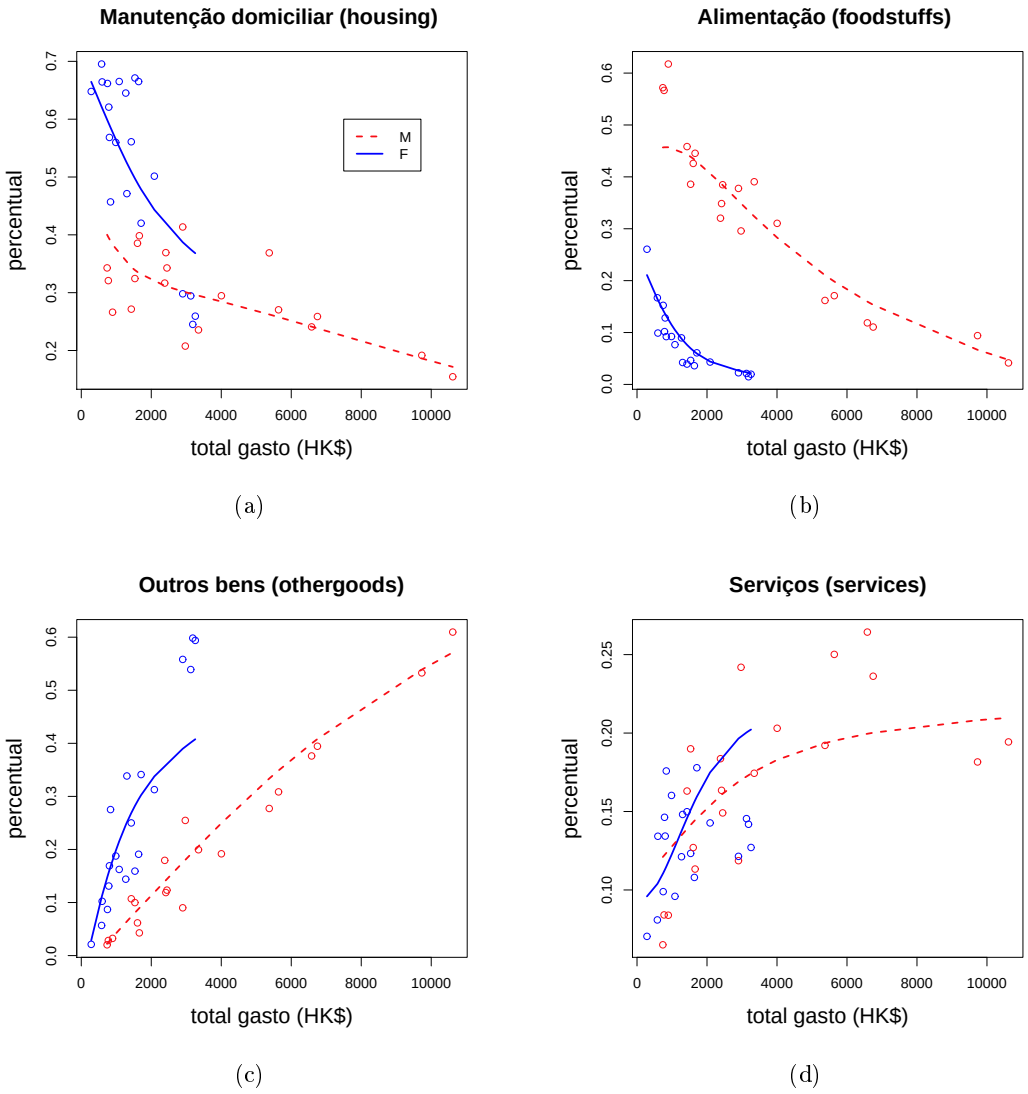


Figura 4.5: *Modelo de Dirichlet linear*

Legenda: $k = 5,\ x_\bullet = (1,\ s,\ r,\ r^2,\ \log(r))$.

$\widehat{\beta} =$									
β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	$1.433e+2$	$-6.138e+0$	$1.149e-2$	$-3.707e-7$	$-1.890e+1$
β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	$8.032e+1$	$-2.561e+1$	$4.677e-3$	$-3.748e-7$	$-8.025e+0$
β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	$-2.695e+1$	$3.571e+0$	$-3.235e-3$	$1.599e-6$	$4.535e+0$
β_{41}	β_{42}	β_{43}	β_{44}	β_{45}	$2.290e+1$	$-4.301e+0$	$2.832e-3$	$2.664e-7$	$-2.580e+0$

Tabela 4.5: *Parâmetros de máxima verossimilhança*

4.2.2 Modelo de Melo, Vasconcellos e Lemonte

A Figura 4.6 mostra o ajuste obtido pelo Modelo de Dirichlet de Melo Vasconcellos e Lemonte correspondente aos parâmetros da máxima verossimilhança da Tabela 4.6. Para esse caso, a função Exponencial não apresentou instabilidade numérica.

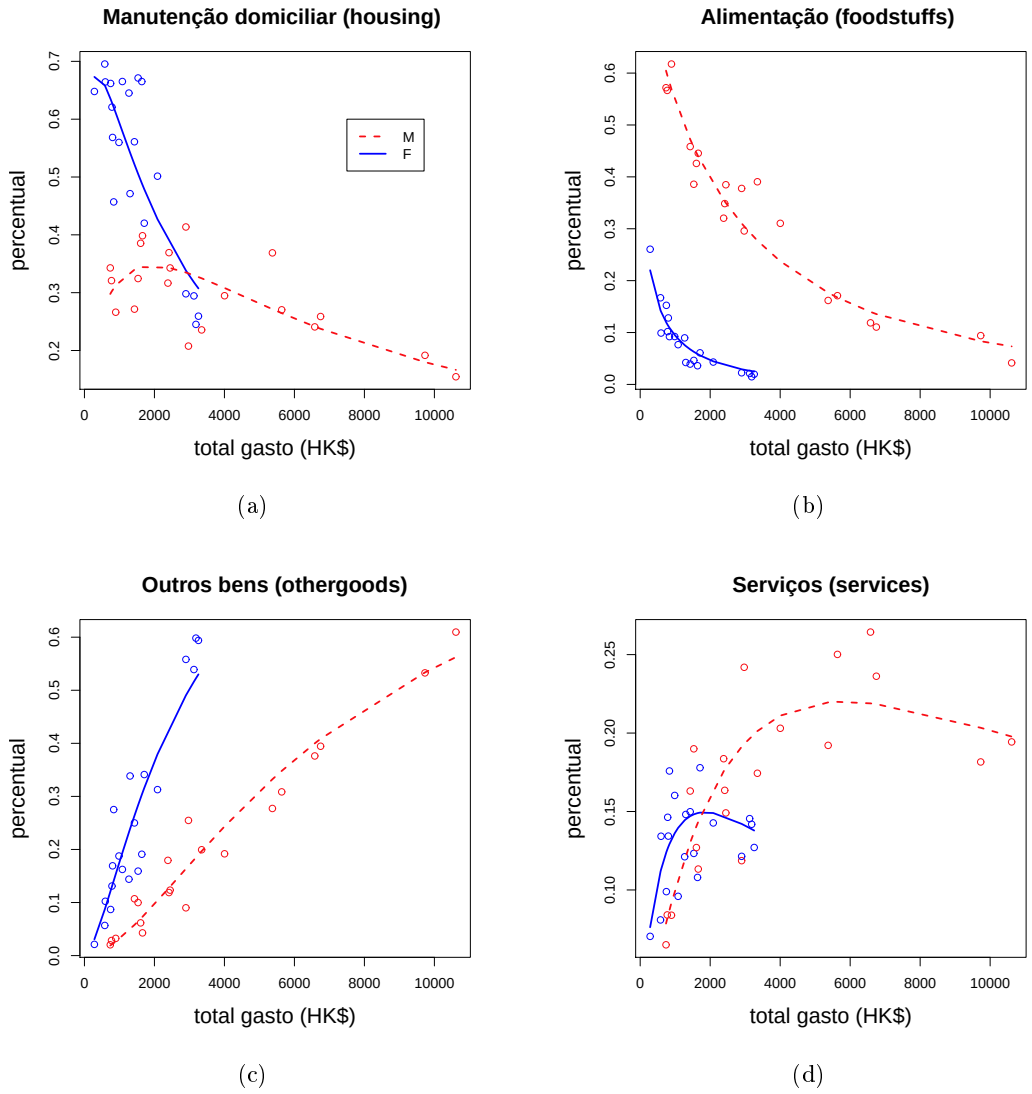


Figura 4.6: *Modelo de Melo, Vasconcellos e Lemonte*

Legenda: $k = 3,\ x_\bullet = (1,\ s,\ log(r)),$
$g_j(u) = exp(u), j = 1, 2, 3.$

$\widehat{\beta} =$					
β_{11}	β_{12}	β_{13}	$6.042e+0$	$-1.078e-2$	$-3.335e-1$
β_{21}	β_{22}	β_{23}	$1.054e+1$	$-2.384e+0$	$-9.069e-1$
β_{31}	β_{32}	β_{33}	$-6.554e+0$	$1.077e+0$	$1.156e+0$
β_{41}	β_{42}	β_{43}	$9.840e-1$	$-3.180e-1$	$2.307e-1$

Tabela 4.6: *Parâmetros de máxima verossimilhança*

4.2.3 Modelo linear $TgRatio$ (3.22)

O modelo $TgRatio$ não apresentou bons resultados para as composições do Lago Ártico. Na Figura 4.7, apresentamos o bom ajuste gerado por esse modelo para exemplificar a sua utilidade.

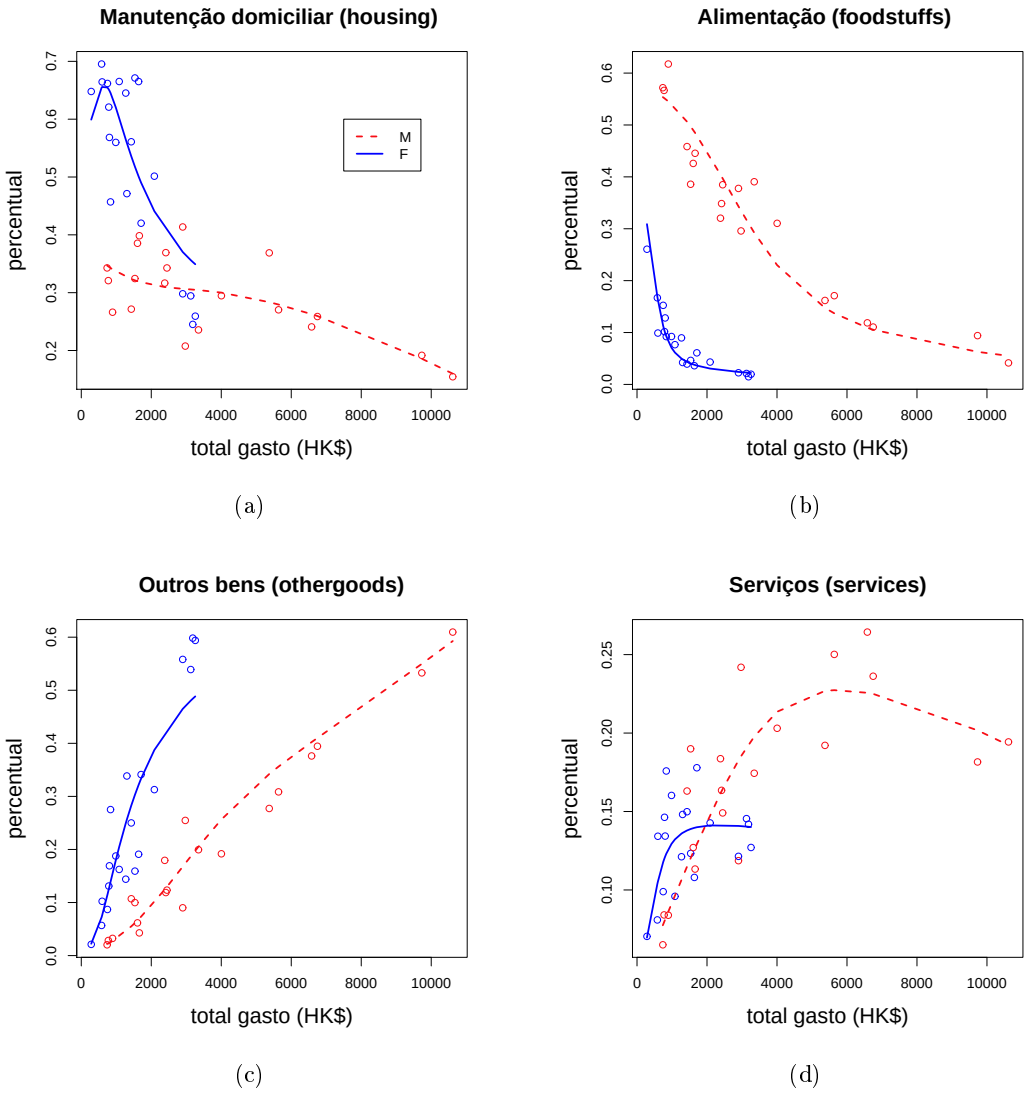


Figura 4.7: *Modelo linear TgRatio*

Legenda: $k = 5,\ x_\bullet = (1,\ z_\bullet,\ z_\bullet^2,\ \log(z_\bullet))$.

$\widehat{\beta} =$									
β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	$2.786e+1$	$1.165e+0$	$1.027e-3$	$-4.407e-8$	$-3.685e+0$
β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	$3.712e+1$	$-7.036e+0$	$3.333-4$	$-1.112e-8$	$-4.597e+0$
β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	$-2.369e+1$	$3.169e+0$	$-7.377e-4$	$4.319e-8$	$3.173e+0$

Tabela 4.7: *Parâmetros de máxima verossimilhança*

4.3 Casamentos por faixa etária (bivariado)

Uma covariável dupla $s = (m, w)$ será utilizada para representar o sexo dos individuos: $s = (1, 0)$ para homens e $s = (0, 1)$ para mulheres. Nos modelos abaixo, utilizamos os dados referentes apenas ao período 1984-1999, para poder comparar as previsões geradas pelos modelos e os dados observados referente aos anos 2000, 2001 e 2002.

4.3.1 Modelo de Dirichlet Linear

A Figura 4.8 mostra o ajuste obtido pelo Modelo de Dirichlet Linear.

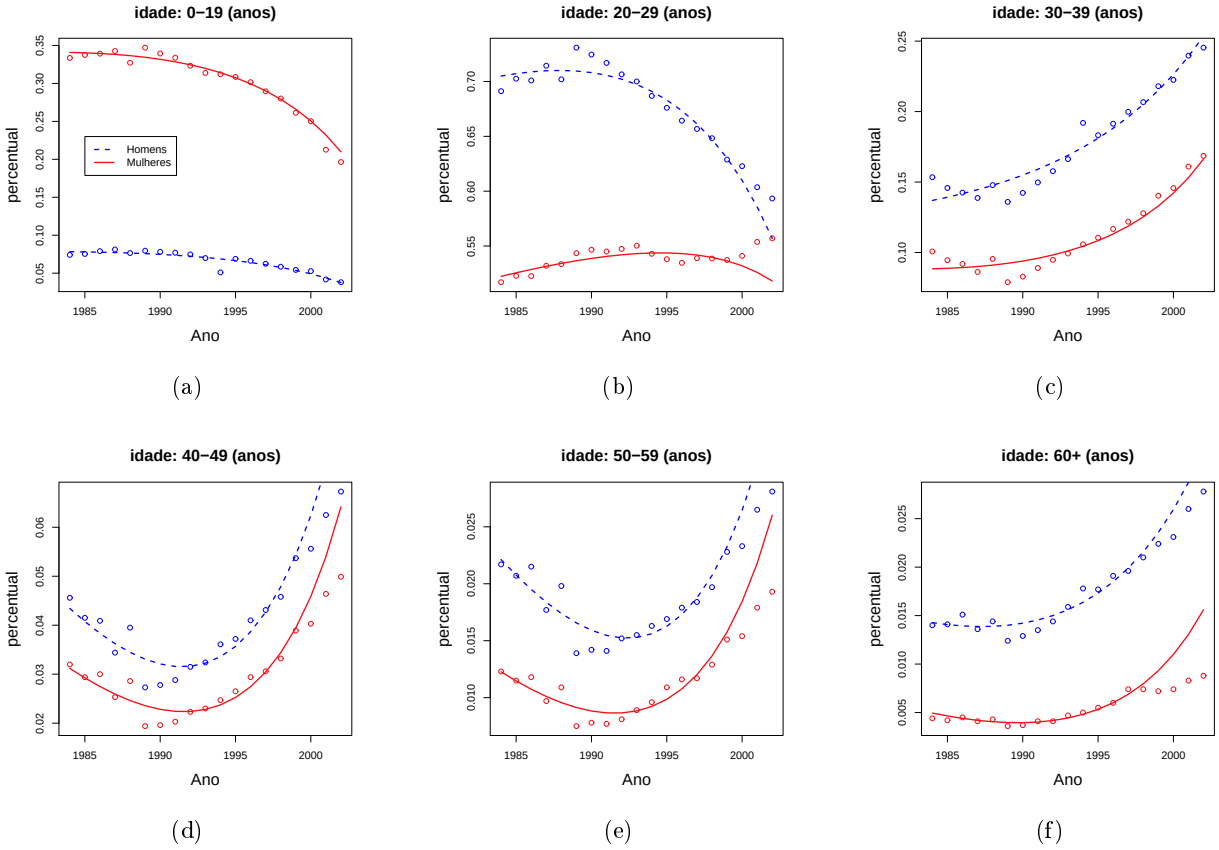


Figura 4.8: *Modelo de Dirichlet Linear*

Legenda:

$$k = 7, \quad x_\bullet = (1,\ w*t,\ w*t^2,\ w*cos(t/100)\ m*t,\ m*t^2,\ m*cos(t/100))$$

$$z_\bullet = (m, w, t)\ s = (m, w) = \text{sexo},\ t = \text{ano}.$$

$\widehat{\beta} =$						
β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}
β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}
β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	β_{36}	β_{37}
β_{41}	β_{42}	β_{43}	β_{44}	β_{45}	β_{46}	β_{47}
β_{51}	β_{52}	β_{53}	β_{54}	β_{55}	β_{56}	β_{57}
β_{61}	β_{62}	β_{63}	β_{64}	β_{65}	β_{66}	β_{67}
$-3.900e+1$	$5.059e+1$	$-2.545e-2$	$2.934e+3$	$2.220e+0$	$-1.199e-3$	$1.165e+3$
$-1.030e+2$	$5.370e+1$	$-2.704e-2$	$4.950e+3$	$5.262e+1$	$-2.656e-2$	$5.098e+3$
$4.964e+2$	$2.441e+2$	$-1.195e-1$	$-2.539e+4$	$1.359e+2$	$-6.657e-2$	$-1.385e+4$
$1.023e+3$	$3.762e+2$	$-1.842e-1$	$-4.037e+4$	$3.716e+2$	$-1.820e-1$	$-3.991e+4$
$4.442e+2$	$1.549e+2$	$-7.584e-2$	$-1.663e+4$	$1.738e+2$	$-8.510e-2$	$-1.859e+4$
$1.822e+2$	$7.204e+1$	$-3.527e-2$	$-7.796e+3$	$6.724e+1$	$-3.292e-2$	$-7.224e+3$

Tabela 4.8: *Parâmetros de máxima verossimilhança*

4.3.2 Modelo de Melo, Vasconcellos e Lemonte

A Figura 4.9 mostra o ajuste obtido pelo Modelo de Dirichlet de Melo, Vasconcellos e Lemonte. A forma das funções positivas g_j, $j = 1, 2, 3, 4, 5, 6$ foi escolhida artificialmente para melhorar a qualidade do ajuste (não conseguimos produzir um ajuste adequado utilizando somente as variantes da função arcotangente, assim como a função Exponencial). O processo de maximização da função de verossimilhança não foi completado, assim como no caso das composições do solo do Lago Ártico.

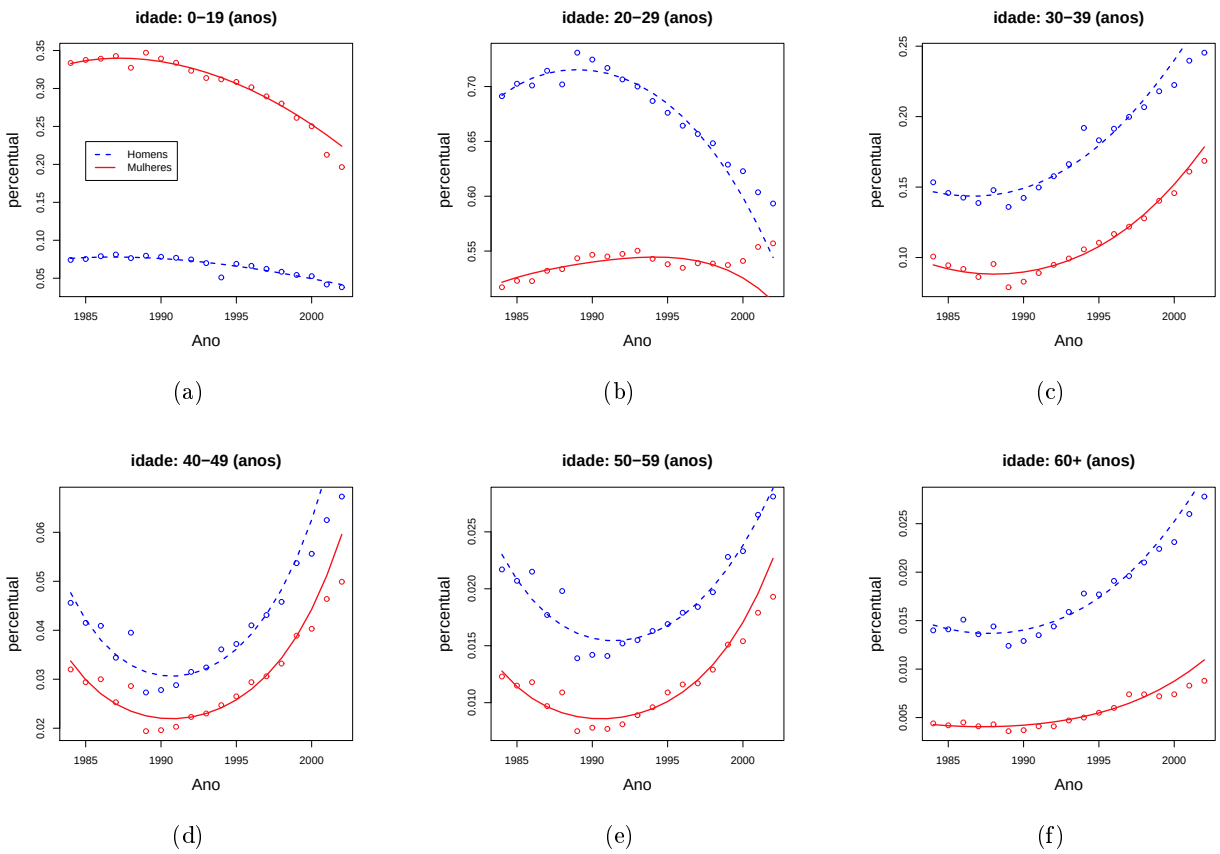


Figura 4.9: *Modelo de Melo, Vasconcellos e Lemonte*

Legenda:

$$k = 7, \quad x_\bullet = (1,\ w*t, w*(t/10)^2, w*cos(t/100),\ m*t,\ m*(t/10)^2,\ m*cos(t/100))$$

$$g_j(u) = (tg^{-1}(u) + 1.6) * exp(u), j = 1,2,3,4,5,6.$$

$$z_\bullet = (m, w, t),\ s = (m, w) = \text{sexo},\ t = \text{ano}.$$

$\widehat{\beta} =$							
		β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}
		β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}
		β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	β_{36}	β_{37}
		β_{41}	β_{42}	β_{43}	β_{44}	β_{45}	β_{46}	β_{47}
		β_{51}	β_{52}	β_{53}	β_{54}	β_{55}	β_{56}	β_{57}
		β_{61}	β_{62}	β_{63}	β_{64}	β_{65}	β_{66}	β_{67}
$-7.347e+0$	$-2.784e+0$	$1.366e-1$	$2.887e+2$	$-2.487e+0$	$1.218e-1$	$2.761e+2$
$-5.121e+0$	$-2.193e+0$	$1.077e-1$	$2.227e+2$	$-1.968e+0$	$9.644e-2$	$2.181e+2$
$-5.902e-1$	$-1.722e-1$	$8.805e-3$	$4.028e-1$	$-5.022e-2$	$2.605e-3$	$5.786e+0$
$7.088e+0$	$2.118e+0$	$-1.034e-1$	$-2.469e+2$	$2.799e+0$	$-1.370e-1$	$-3.019e+2$
$5.340e+0$	$1.947e+0$	$-9.501e-2$	$-2.282e+2$	$1.722e+0$	$-8.426e-2$	$-1.837e+2$
$1.469e+0$	$3.917e-1$	$-1.887e-2$	$-6.211e+1$	$5.062e-1$	$-2.470e-2$	$-5.524e+1$

Tabela 4.9: *Parâmetros de máxima verossimilhança*

4.3.3 Modelo linar Log-esférico

A Figura 4.10 mostra o ajuste obtido pelo Modelo Log-esférico, permutando-se as componentes das composições de forma conveniente.

Legenda: $T = T_{logEsf} \circ P, \quad P(y_1, y_2, y_3, y_4, y_5, y_6) = (y_3, y_4, y_1, y_2, y_5, y_6),$

$$k = 6, \qquad x_\bullet = (w,\ w*t,\ w*t^4, m,\ m*t,\ m*t^4),$$

$$z_\bullet = (m, w, t),\ s = (m, w) = \text{sexo},\ t = \text{ano}.$$

$\widehat{\beta} =$					
		β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}
		β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}
		β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	β_{36}
		β_{41}	β_{42}	β_{43}	β_{44}	β_{45}	β_{46}
		β_{51}	β_{52}	β_{53}	β_{54}	β_{55}	β_{56}
$2.203e+3$	$-1.480e+0$	$4.739e-11$	$1.614e+3$	$-1.084e+0$	$3.484e-11$
$1.094e+4$	$-7.335e+0$	$2.330e-10$	$8.200e+3$	$-5.503e+0$	$1.753e-10$
$1.947e+4$	$-1.304e+1$	$4.131e-10$	$1.958e+4$	$-1.311e+1$	$4.156e-10$
$1.976e+4$	$-1.324e+1$	$4.194e-10$	$1.683e+4$	$-1.127e+1$	$3.569e-10$
$1.384e+4$	$-9.287e+0$	$2.953e-10$	$1.020e+4$	$-6.845e+0$	$2.178e-10$

Tabela 4.10: *Parâmetros de máxima verossimilhança*

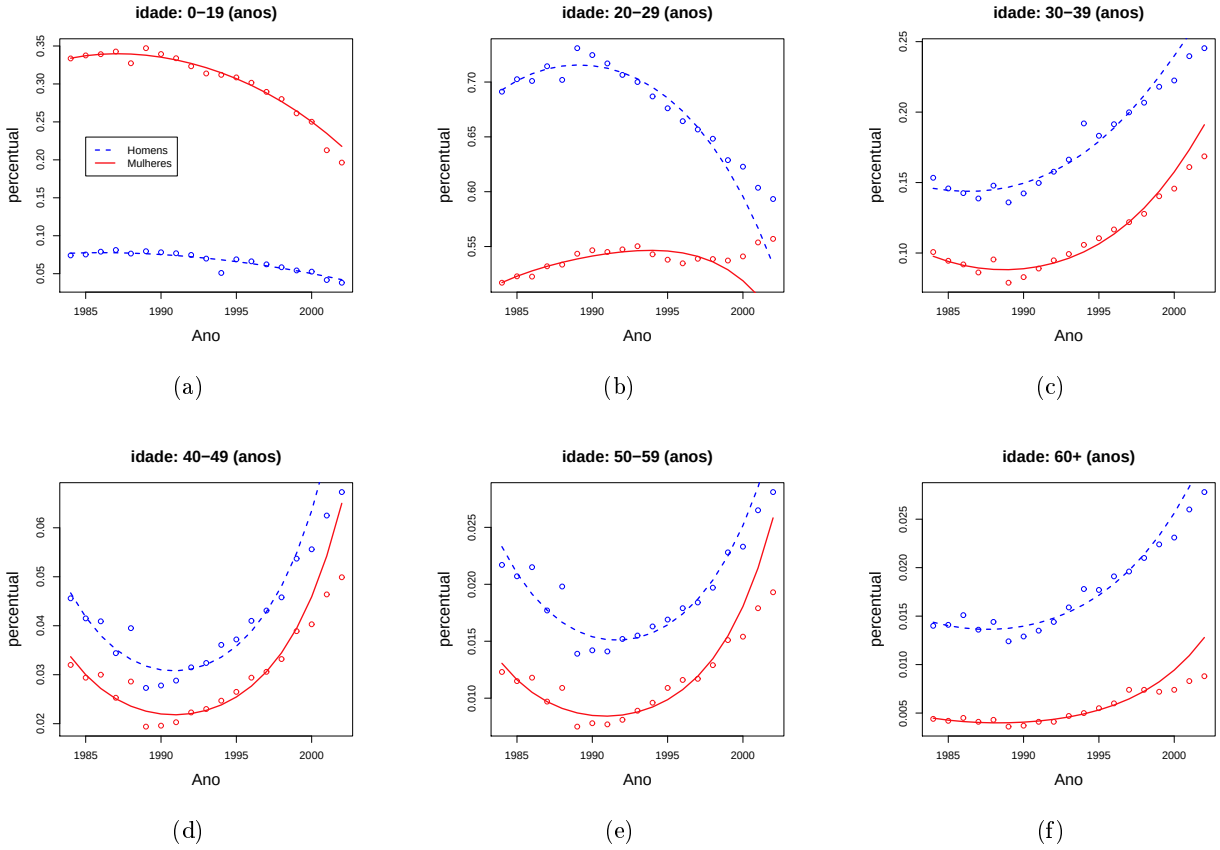


Figura 4.10: *Modelo Log-esférico*

4.4 Considerações sobre o Espaço Paramétrico

No Capítulo introdutório, foi apresentado um método geral para a obtenção de modelos de regressão, onde a função preditora $Prev$ associa, a cada valor da variável explicativa $x_\bullet$, o valor esperado da distribuição com parâmetros $\psi(x_\bullet, \beta)$. Em alguns casos, pode-se querer calcular algumas medidas de dispersão em torno dos valores ajustados, o que pode ser feito calculando (quando possível) os momentos de ordens superiores da distribuição $h(.|\psi(x_\bullet, \beta))$. Dessa forma, o funcional que associa, a cada vetor/matriz de parâmetros β, a função $\psi(\ .\ , \beta)$ é mais completo do que o funcional que associa β à função $Prev$. Porém, se consideramos uma igualdade entre modelos que possuem funções de previsão idênticas, então é preciso reavaliar o método de obtenção da função $Prev$ para alguns dos modelos propostos, pois estes encontram-se descaracterizados.

4.4.1 Modelo de Dirichlet/ Mínimos quadrados (linear)

Definição: Dizemos que um subconjunto fechado de $\mathbb{R}^m$, K, é um cone se:

- $u + v \in\ K\ \forall\ u, v \in K$.
- $\alpha u\ \in K,\ \forall\, u \in K$ e $\alpha \geq 0$.

O cone K é dito próprio se $K \cap -K = \{0\}$, e sólido, se $K/\partial K \neq \emptyset$.

Em (2.2), focamos em modelos da forma

$$g = \left(\frac{\lambda_1}{\sum \lambda_j}, \frac{\lambda_2}{\sum \lambda_j}, \ldots, \frac{\lambda_D}{\sum \lambda_j} \right), \tag{4.1}$$

onde $\lambda_1(X), \lambda_2(X), \ldots, \lambda_D(X)$ são dadas por (2.11).

Em termos (2.4), o espaço paramétrico Θ é constituído por todas as matrizes β (ou vetores em $\mathbb{R}^{r*k}$) tais que $\beta \times x_{j\bullet} > 0,\ j = 1, 2, \ldots, n$. Temos, então, que $\overline{\Theta} = \Theta \cup \partial\Theta$ é um cone próprio e sólido de $\mathbb{R}^{r*k}$. Porém, devido à forma de (4.1), ao multiplicarmos β por uma constante positiva, a função preditora g permanece a mesma. Isso, de certa forma, descaracteriza a escolha do estimador de máxima verossimilhança, pois a função de verossimilhança não é invariante sobre dilatações dos seus argumentos e, portanto, existem pontos do espaço paramétrico (cuja verossimilhança é muito baixa) que produzem os mesmos resultados de pontos com verossimilhança alta. O gráfico 4.11 mostra o comportamento da função do logaritmo da verossimilhança (2.8) ao dilatar/contrair os parâmetros do modelo quadrático MDQ da tabela 4.7.

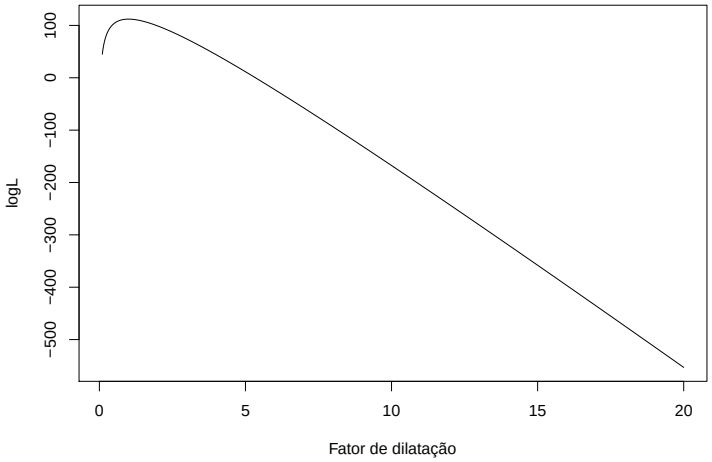


Figura 4.11: *Sensibilidade da verossimilhança de MDQ sobre dilatações*

Da mesma forma, a soma dos resíduos quadráticos (3.26) é invariante sobre dilatações de β, para h como em (4.1), de modo que o estimador de mínimos quadrados também fica descaracterizado. Uma forma de eliminar essa ambiguidade é considerar o espaço paramétrico $\Theta' = \{\theta \in \Theta : ||\theta|| = 1\}$. Utilizando as coordenadas esféricas (3.18), o novo espaço paramétrico fica reduzido a um subconjunto de $[0, 2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{k*r-2}$. Além de "enxugar o excesso de pontos"do espaço paramétrico, trabalhar sobre Θ' (um conjunto limitado) é mais vantajoso, em termos de integração (exata ou numérica), do que trabalhar sobre um conjunto ilimitado. Por outro lado, maximizar a função de verossimilhança nas variáveis $\alpha_1, \alpha_2, \ldots \alpha_{k*r-1}$ de (3.18) é deveras mais trabalhoso, pois as restrições sobre os novos parâmetros não são mais lineares, como no outro caso. Para o caso do Lago Ártico, conseguimos tanto maximizar a verossimilhança reparametrizada, assim como minimizar a soma dos resíduos quadráticos, redefinindo as funções e os seus gradientes (a valores bem baixos) na região de $[0, 2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{k*r-2}$ não correspondente à Θ'. Esse procedimento, porém, produz

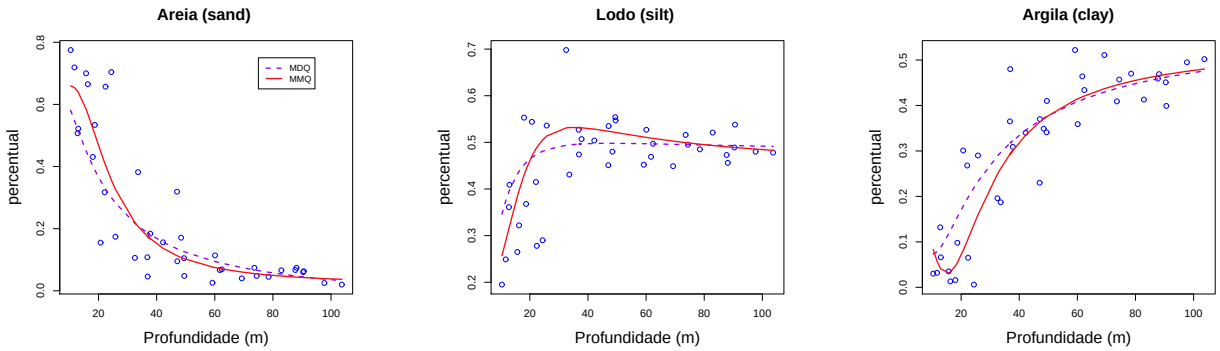
descontinuidades que geram imcompatibilidades na redefinição dos gradientes, de modo que essa prática não é aconselhavel.

$$
\begin{array}{lcl}
\beta_{11} & = & \cos(\alpha_1)\cos(\alpha_2)\cos(\alpha_3)\cos(\alpha_4)\cos(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{12} & = & sen(\alpha_1)\cos(\alpha_2)\cos(\alpha_3)\cos(\alpha_4)\cos(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{13} & = & sen(\alpha_2)\cos(\alpha_3)\cos(\alpha_4)\cos(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{21} & = & sen(\alpha_3)\cos(\alpha_4)\cos(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{22} & = & sen(\alpha_4)\cos(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{23} & = & sen(\alpha_5)\cos(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{31} & = & sen(\alpha_6)\cos(\alpha_7)\cos(\alpha_8) \\
\beta_{32} & = & sen(\alpha_7)\cos(\alpha_8) \\
\beta_{33} & = & sen(\alpha_8)
\end{array}
$$

Tabela 4.11: *Transformação de MDQ para coordenadas esféricas*

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
MDQ	2.152e-1	-1.291e-3	-1.200e-1	5.699e-2	1.232e-2	9.899e-1	-1.441e-1	8.068e-3
MMQ	0.000e+0	1.320e-4	3.454e-1	-2.780e-2	2.985e-3	7.101e-1	-8.292e-2	2.800e-3

Tabela 4.12: *Parâmetros de máxima verossimilhança/ Mínimos quadrados*



(a) (b) (c)

Figura 4.12: *Modelos quadráticos restritos à casca esférica unitária*

Capítulo 5

Seleção de modelos

Como vimos na Seção 4.4, o "excesso" de parâmetros em alguns modelos de regressão pode trazer alguns inconvenientes como, por exemplo, a perda da unicidade do estimador de mínimos quadrados. Além disso, como relatado em Melo *et al.* (2009), a sensibilidade dos estimadores à perturbações pode aumentar consideravelmente com o aumento do número de parâmetros preditores. Por outro lado, modelos simples demais (com uma quantidade muito pequena de parâmetros) podem comprometer o ajuste ao conjunto de dados observados. A obtenção de um equilíbrio entre esses dois extremos, segundo Draper e Smith (1998), pág. 294, constitui o que comumente é designado por "selecionar o melhor modelo de regresssão". Um outro problema de fundamental importância relacionado ao de seleção de modelos é o da análise da dependência da variável resposta em função das variáveis explicativas (covariáveis), visto que, em alguns modelos, essa dependência se traduz na presença/ausência de certos parâmetros. Neste capítulo apresentamos alguns métodos de seleção de modelos baseados em inferência Bayesiana.

5.1 BIC

O BIC (*Bayesian Information Criterion*) introduzido por Schwarz (1978) fornece um método relativamente simples para selecionar, dentre um conjunto finito de modelos, o modelo mais simples e representativo para um conjunto de dados observados $Y = \{y_1, y_2, \ldots, y_n\}$.

Particularidades (segundo Schwarz):

1. O espaço paramétrico Θ deve ser um subconjunto convexo de $\mathbb{R}^m$.

2. Assume-se que a variável y possua distribuição $f(y|\theta) = \exp\{T(y) \times \theta - b(\theta)\}$, pertencente à familia exponencial, para algum $\theta \in \Theta$.

3. Os modelos candidatos são da forma $\Theta_j = v_j \cap \Theta$, onde v_j é um subespaço k_j-dimensional de $\mathbb{R}^m$ $(j = 1, 2, \ldots, s)$.

Schwarz assume *prioris* do tipo $\sum_{j=1}^{s} \alpha_j \overline{g_j}$, onde $g_1, g_2, \ldots, g_s$ são funções densidade de probabilidade definidas respectivamente em $\Theta_1, \Theta_2, \ldots, \Theta_s$, $\alpha_1, \alpha_2, \ldots, \alpha_s$ são números não negativos tais que $\sum_{j=1}^{s} \alpha_j = 1$ e $\overline{g_j}(\Omega) = g_j(\Omega \cap \Theta_j)$, $\forall\ \Omega \subset \Theta$. A solução ótima corresponde a escolher o modelo mais provável a *posteriori*, isto é: devemos escolher j de modo a maximizar

$$S(Y, n, j) = \log \int_{\Theta_j} \left(\alpha_j \exp\left\{ n \left(\widetilde{T}(Y) \times \theta - b(\theta) \right) \right\} \right) dg_j^*(\theta), \tag{5.1}$$

onde $\widetilde{T}(Y) = \frac{1}{n} \sum_{i=1}^{n} T(y_i)$.

O método, que consiste em escolher o modelo que minimiza[1] a quantidade

$$BIC = -\max\{\log L(\theta|Y) : \theta \in \Theta_j\} + \tfrac{1}{2} k_j \log(n),$$

baseia-se na proposição que afirma que

$$S(Y, n, j) = n\, sup \left\{ \left(\widetilde{T}(Y) \times \theta - b(\theta) \right) : \theta \in \Theta_j' \right\} - \frac{1}{2} k_j \log(n) + R \tag{5.2}$$

onde R é uma função limitada em n, para Y e j fixados.

A prova é separada em 3 lemas:

Lema 1. *A fórmula* (5.2) *vale se* $\widetilde{T}(Y) \times \theta - b(\theta)$ *é da forma* $A - \lambda ||\theta - \theta_0||^2$, *para números positivos* A, λ *e* $\theta_0 \in \Theta$ *fixado.*

Lema 2. *Se as funções densidade de probabilidade de duas variáveis aleatórias limitadas e positivas* U *e* V *coincidem onde ambas excedem* ρ, *para algum* ρ, $0 < \rho < \sup U$, *então*

$$\lim_{n \to \infty} \log E(U^n) = \lim_{n \to \infty} \log E(V^n)$$

Lema 3. *Exitem constantes* $\lambda_1, \lambda_2, \rho > 0$ *tais que*

$$A - \lambda_1 ||\theta - \theta_j||^2 < \widetilde{T}(Y) \times \theta - b(\theta) < A - \lambda_2 ||\theta - \theta_j||^2 \tag{5.3}$$

sempre que $\exp\left\{ \widetilde{T}(Y) \times \theta - b(\theta) \right\} > \rho$, *onde* $A = \sup \widetilde{T}(Y) \times \theta - b(\theta)$ *e* $0 < \rho < e^A$.

A fórmula (5.2) nos diz que, assintoticamente, o termo limitado pode ser descartado, o que leva à escolha do modelo com o menor BIC.

[1] Originalmente, o critério de Schwarz propõe escolher o modelo que maximiza $-BIC$.

A proposição de Schwarz é obtida combinando-se os Lemas 1, 2 e 3. O argumento é analogo ao "Teorema do Sanduiche"' para limites: no Lema 1, prova-se a proposição assumindo que a função de verossimilhança possui uma forma específica. O Lema 3 noz diz que, numa vizinhança do ponto de máxima verossimilhança, a função de verossimilhança pode ser prençada entre duas funções para as quais a proposição é válida (pelo Lema 1) e o Lema 2 conclui que, sob essas condições, o resultado vale para a função de verossimilhança.

Embora bastante utilizado em problemas de seleção de modelos, a base matemática apresentada por Schwarz para a sustentação do BIC apresenta algumas inconsistências. Note que, em geral, Θ_j é um subconjunto de medida (de Lebesque) nula em Θ. Segundo os fundamentos da teoria bayesiana ortodoxa, hipótese precisas (hipóteses envolvendo subconjuntos de medida nula do espaço paramétrico) não fazem sentido (Stern (2008), pág 29), pois conjuntos de medida nula possuem probabilidade ZERO. Por conseguinte, nas diversas justificativas criadas para contornar esse tipo de impasse, frequentemente podem ser encontrados argumentos passíveis de erros de interpretação e aparentemente contraditórios (Schwarz admite *prioris* que atribuem probabilidade não nula aos conjuntos Θ_js). Além disso, outros argumentos paradoxais são apresentados ao longo do texto:

1. Observe, em primeiro lugar, que a medida g_j^* é a medida induzida pela função g_j, isto é:

$$g_j^*(\Omega) = \int_{\Omega} g_j(\theta) d\theta \quad \forall\, \Omega \subset \Theta. \tag{5.4}$$

   No Lema 1, Schwarz assume que g_j^* é a medida de Lebesgue em v_j, isto é: $g_j^*(\Omega) = \int_{\Omega} d\theta$ para todo Ω mensurável. Isso, obviamente, só é possível se $g_j \equiv 1$ (assumindo-se que g_j seja, ao menos, contínua). Daí segue que

$$\int_{\Theta_j} d\theta = 1, \tag{5.5}$$

   pois g_j é uma função densidade de probabilidade. Essa igualdade, porém, não é satisfeita para conjuntos convexos Θ arbitrários, como suposto pelo autor.

2. Mesmo assumindo a hipótese anterior (o que já nos leva a uma contradição) não há nenhuma outra proposição apresentada pelo autor que estabeleça uma relação deste com o caso geral.

3. Ainda no Lema 1, supondo que $\widetilde{T}(Y) \times \theta - b(\theta)$ é da forma $A - \lambda||\theta - \theta_0||^2$, para números positivos A, λ e $\theta_0 \in \Theta_j$ fixado, o autor exibe explicitamente o cálculo da integral (5.1)

$$\int_{\Theta_j} \left(\alpha_j \exp\left\{n(A - \lambda||\theta - \theta_0||^2)\right\}\right) d\theta = \alpha_j (\pi/n\lambda)^{k_j/2} e^{nA}. \tag{5.6}$$

   Novamente, a igualdade (5.6) não é satisfeita para conjuntos convexos $\Theta_j = \Theta \cap v_j$ em geral.

Para $\Theta_j = \mathbb{R}^{k_j}$, a fórmula é verdadeira (bem familiar aos introduzidos à distribuição Normal multivariada). Porém, essa hipótese e (5.5) também são contraditórias.

4. É importante ressaltar que, se $\Theta_i \subset \Theta_j$ propriamente, isto é, $v_i \subset v_j$ e $\dim v_i < \dim v_j$, então o uso de medidas distintas em Θ_i e Θ_j é contraditório. Pois, se assumimos a medida de Lebesgue em Θ_j, então todo subconjunto de Θ_i tem medida nula em Θ_j. Logo, o uso da medida de Lebesgue também em Θ_i é incoerente, pois a probabilidade de Θ_i não pode ser nula e não-nula simultaneamente. Daí conclui-se que o critério de Schwarz faz sentido somente para comparar modelos Θ_i e Θ_j tais que

$$\Theta_i \not\subset \Theta_j. \tag{5.7}$$

Isso, porém, talvez já estivesse implícito nas palavras de Schwarz: "the competing models are...", uma vez que não faz sentido concorrer um modelo com um de seus submodelos.

A seguir, apresentamos uma possível formalização dessas idéias.

5.1.1 Uma proposta de formalização do BIC

Considere o espaço paramétrico $\Theta' = \bigcup_{j=1}^{s} \Theta_j$. Se $\mu_1, \mu_2, \ldots, \mu_s$ representam as medidas de Lebesgue em $\Theta_1, \Theta_2, \ldots, \Theta_s$, respectivamente, então μ', dada por $\mu'(\Omega) = \sum_{j=1}^{s} \mu_j(\Omega \cap \Theta_j)$, é uma medida positiva em Θ'. Com essa nova medida, se $g_1, g_2, \ldots, g_s$ são funções densidade de probabilidade (com relação à medida de Lebesgue) definidas, respectivamente, em $\Theta_1, \Theta_2, \ldots, \Theta_s$ e $\alpha_1, \alpha_2, \ldots, \alpha_s$ são números não negativos tais que $\sum_{j=1}^{s} \alpha_j = 1$, então $\sum_{j=1}^{s} \alpha_j \overline{g_j}$ representa uma função densidade de probabilidade em (Θ', μ') ($\overline{g_j}(\Omega) = g_j(\Omega \cap \Theta_j)$). Reciprocamente, toda função densidade de probabilidade h (contínua no sentido usual), definida em (Θ', μ'), pode ser posta nesse formato, fazendo $\alpha_j = \frac{\int_{\Theta_j} h\, d\mu'}{\int_{\Theta'} h\, d\mu'}$ e $g_j = \frac{1}{\alpha_j} h|_{\Theta_j}$, pois[2] $\mu'|_{\Theta_j} = \mu_j$.

Portanto, podemos assumir, sem perda de generalidade, que a distribuição a *posteriori* em (Θ', μ') é da forma

$$\left(\sum_{j=1}^{s} \alpha_j g_j(\theta) \right) \exp\left\{ n \left(\widetilde{T}(Y) \times \theta - b(\theta) \right) \right\}, \tag{5.8}$$

onde $\widetilde{T}(Y) = \frac{1}{n} \sum_{i=1}^{n} T(y_i)$.

[2] Por (5.7), se $\Omega \subset \Theta_j$, então $\mu_i(\Omega) = \mu_i(\Omega \cap \Theta_i \cap \Theta_j) = 0$, pois $v_i \cap v_j$ possui medida mula em Θ_i, para $i \neq j$.

Como mencionado anteriormente, a estratégia de Schwarz consiste em escolher o modelo que maximiza

$$S(Y,n,j) = \log \int_{\Theta_j} \left(\alpha_j g_j(\theta) \exp\left\{ n \left(\widetilde{T}(Y) \times \theta - b(\theta) \right) \right\} \right) d\mu'(\theta). \tag{5.9}$$

Suponha que as integrais em (5.9) são definidas sobre subconjuntos da forma

$$\Theta'_j = v_j \cap \{\theta \in \mathbb{R}^m : ||\theta - \theta_j||^2 \leq r_j\}, \tag{5.10}$$

onde $\int_{\Theta'} d\mu_j = 1$ e θ_j é o ponto de máximo da verossimilhança restrita à v_j. Supondo que, de fato, um (e somente um) ponto do espaço paramétrico corresponde aos dados observados e que o estimador de máxima verossimilhança é consistente (Pawitan (2001), pág. 256), a restrição dos modelos aos conjuntos descritos em (5.10), assim como a escolha da priori, não deve alterar os resultados assintóticos. Portanto, podemos supor que a priori é constante e igual a 1 em cada um dos Θ'_js (obviamente, definimos os Θ'_js de modo a possibilitar uma tal escolha).

As provas dos lemas 2 e 3 dadas por Schwarz são bem acessíveis e não apresentam nenhuma inconsistência, adaptando-se facilmente assumindo (5.10) (recomendamos a leitura de Schwarz (1978) para maiores detalhes). A seguir provamos o Lema 1 sob essas novas hipóteses para obtermos o resultado de Schwarz.

Prova do Lema 1:

Por indução em $k \geq 2$, defina o sistema de coordenadas:

$$\left\{ \begin{array}{lcl} x_1^{(2)} & = & \rho\cos(\beta_1) \\ x_2^{(2)} & = & \rho sen(\beta_1) \end{array} \right. \quad e \quad \left\{ \begin{array}{lcl} x_1^{(k+1)} & = & x_1^{(k)}\cos(\beta_k) \\ & \vdots & \\ x_k^{(k+1)} & = & x_k^{(k)}\cos(\beta_k) \\ x_{k+1}^{(k+1)} & = & \rho sen(\beta_k) \end{array} \right. , \quad \beta_i \in \left\{ \begin{array}{ll} [0, 2\pi], & \text{se } i = 1 \\ \left[-\frac{pi}{2}, \frac{\pi}{2}\right], & \text{se } i > 1 \end{array} \right. \tag{5.11}$$

O jacobiano é dado por:

$$\frac{\partial(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{k+1}^{(k+1)})}{\partial(\rho, \beta_1, \ldots, \beta_k)} = \left| \begin{array}{ccccc} \frac{\partial x_1^{(k+1)}}{\partial \rho} & \frac{\partial x_1^{(k+1)}}{\partial \beta_1} & \cdots & \frac{\partial x_1^{(k+1)}}{\partial \beta_{k-1}} & \frac{\partial x_1^{(k+1)}}{\partial \beta_k} \\ \frac{\partial x_2^{(k+1)}}{\partial \rho} & \frac{\partial x_2^{(k+1)}}{\partial \beta_1} & \cdots & \frac{\partial x_2^{(k+1)}}{\partial \beta_{k-1}} & \frac{\partial x_2^{(k+1)}}{\partial \beta_k} \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial x_k^{(k+1)}}{\partial \rho} & \frac{\partial x_k^{(k+1)}}{\partial \beta_1} & \cdots & \frac{\partial x_k^{(k+1)}}{\partial \beta_{k-1}} & \frac{\partial x_k^{(k+1)}}{\partial \beta_k} \\ \frac{\partial x_{k+1}^{(k+1)}}{\partial \rho} & \frac{\partial x_{k+1}^{(k+1)}}{\partial \beta_1} & \cdots & \frac{\partial x_{k+1}^{(k+1)}}{\partial \beta_{k-1}} & \frac{\partial x_{k+1}^{(k+1)}}{\partial \beta_k} \end{array} \right| =$$

$$\begin{vmatrix} \frac{\partial x_1^{(k)}}{\partial \rho}\cos(\beta_k) & \frac{\partial x_1^{(k)}}{\partial \beta_1}\cos(\beta_k) & \dots & \frac{\partial x_1^{(k)}}{\partial \beta_{k-1}}\cos(\beta_k) & -x_1^{(k)} sen(\beta_k) \\ \frac{\partial x_2^{(k)}}{\partial \rho}\cos(\beta_k) & \frac{\partial x_2^{(k)}}{\partial \beta_1}\cos(\beta_k) & \dots & \frac{\partial x_2^{(k)}}{\partial \beta_{k-1}}\cos(\beta_k) & -x_2^{(k)} sen(\beta_k) \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial x_k^{(k)}}{\partial \rho}\cos(\beta_k) & \frac{\partial x_k^{(k)}}{\partial \beta_1}\cos(\beta_k) & \dots & \frac{\partial x_k^{(k)}}{\partial \beta_{k-1}}\cos(\beta_k) & -x_k^{(k)} sen(\beta_k) \\ sen(\beta_k) & 0 & \dots & 0 & \rho\cos(\beta_k) \end{vmatrix} \tag{5.12}$$

Por (3.18), vale que $x_i^{(k)} = \rho \frac{\partial x_i^{(k)}}{\partial \rho}$. Então, desenvolvendo (5.12) por Laplace com relação à última linha, obtemos:

$$\frac{\partial(x_1^{(k+1)}, x_2^{(k+1)}, \ \dots, \ x_{k+1}^{(k+1)})}{\partial(\rho, \ \beta_1, \ \dots, \ \beta_k)} = \rho\cos(\beta_k)^{k-1} \ \frac{\partial(x_1^{(k)}, x_2^{(k)}, \ \dots, \ x_k^{(k)})}{\partial(\rho, \ \beta_1, \ \dots, \ \beta_{k-1})} \tag{5.13}$$

Segue que

$$\frac{\partial(x_1^{(k)}, x_2^{(k)}, \ \dots, \ x_k^{(k)})}{\partial(\rho, \ \beta_1, \ \dots, \ \beta_{k-1})} \ = \ \rho^{k-1} \prod_{i=2}^{k} \cos(\beta_{i-1})^{i-2} \tag{5.14}$$

Portanto, pelo teorema de mudança de variáveis na integral, obtemos $S(Y, n, j) =$

$$\log\left[\alpha_j e^{nA} \int_0^{r_j}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} \dots \int_{-\pi/2}^{\pi/2} e^{-n\lambda\rho^2} \rho^{k_j-1} \prod_{i=2}^{k_j} \cos(\beta_{i-1})^{i-2} \ d\beta_{k_j-1} \ \dots \ d\beta_2 \ d\beta_1 \ d\rho\right] \tag{5.15}$$

Fazendo $u = n\rho^2$ e $M = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} \dots \int_{-\pi/2}^{\pi/2} \left(\prod_{i=2}^{k_j} \cos(\beta_{i-1})^{i-2}\right) \ d\beta_{k_j-1} \ \dots \ d\beta_2 \ d\beta_1$, então:

$$S(Y, n, j) = nA + -\frac{k_j}{2}\log(n) + \overbrace{\log\left[\alpha_j \frac{M}{2} \int_0^{n\, r_j^2} e^{-\lambda u} \, u^{\frac{k_j}{2}-1} \ d\, u\right]}^{R} \tag{5.16}$$

R é limitada em n, pois $0 \ < \ \int_0^{n\, r_j^2} e^{-\lambda u} \, u^{\frac{k_j}{2}-1} \ d\, u \ \leq \ \int_0^{\infty} e^{-\lambda u} \, u^{\frac{k_j}{2}-1} \ d\, u \ < \ \infty$.

□

5.2 FBST

Um método elegante e rigoroso para tratar hipóteses precisas, o *Full Bayesian Significant Test* (FBST), foi apresentado em Pereira e Stern (1999), onde foi introduzido o conceito de valor epistêmico (e-valor) de uma hipótese precisa. O e-valor (ev) é uma medida de crença desenvolvida especialmente (embora não exclusivamente) para avaliar a plausibilidade de uma hipótese precisa sobre o espaço paramétrico.

5.2.1 Definição formal

Considere uma família paramétrica de modelos $\{f(y|\theta) : \theta \in \Theta\}$. Seja $H_0 : \theta \in \Theta_{H_0}$ uma hipótese precisa[3] sobre θ e $p_n(\theta|Y)$ a função densidade de probabilidade a *posteriori*, dada pela regra de Bayes, após a observação de uma amostra $Y = \{y_1, y_2, \ldots, y_n\} \subset \Omega$. Inicialmente, o e-valor foi definido da seguinte forma:

$$ev(H_0) = 1 - \int_{T(H_0)} p_n(\theta|Y)d\theta, \tag{5.17}$$

onde $T(H_0) = \{\theta \in \Theta : p_n(\theta|y) > \theta_0\}$, $\theta_0 = arg \max_{\theta \in H_0} p_n(\theta|Y)$. Posteriormente, essa definição foi ligeiramente modificada (Madruga *et al.* (2003)) para tornar o e-valor invariante sobre mudanças de coordenadas no espaço paramétrico, definindo $\theta_0 = arg \max_{\theta \in H_0} \frac{p_n(\theta|Y)}{r(\theta)}$, onde r é uma função densidade de probabilidade de referência escolhida adequadamente (uma para cada teste). A função $s(\theta) = \frac{p_n(\theta|Y)}{r(\theta)}$ é conhecida como a surpresa a *posteriori* (*posteriori surprise*) em relação à função de referência r. O conceito de surpresa, discutido em detalhes em Evans (1997), está relacionado à idéia de que "a ocorrência de um valor x_0 é surpreendente se $f(x_0) = P(x_0)$ é pequeno quando comparado a todos os outros valores possíveis[4] de $f(x)$ ". Uma maneira de medir a surpresa de um valor x_0 é comparar o aumento/diminuição da probabillidade a *posteriori* de x_0 em relação à probabilidade à priori de x_0. Isso sugere utilizar uma priori não-informativa como função de referência.

Análogo aos métodos clássicos de inferência, o FBST consiste em designar uma região crítica de rejeição à hipótese: $RC = \{Y \subset \Omega : ev(H_0|Y) < \overline{e}\}$, onde $\overline{e}$ é um limiar fixado a critério do observador. Porém, determinar o quão grande/pequeno deve ser tal limiar está longe de ser uma questão trivial. Uma alternativa a esse impasse, quando disponível, é o uso do $p-$valor (ou nível de significância) (DeGroot e Schervish (2002), pág. 455). Pereira *et al.* (2008) apresentam uma aproximação do $p - valor$ para o FBST utilizando o Teorema da Aproximação Normal para o estimador de máxima verossimilhança (Schervish (1997), página 421), juntamente com a prova da consistência do $e-$valor, sob essas hipóteses.

[3] Assume-se que Θ_{H_0} é uma variedade com dimensão menor do que a dimensão de Θ.
[4] Evans se refere a um modelo discreto, nessa ocasião.

5.3 Teste da Razão de verossimilhança (LR)

Um dos testes mais conhecidos, o teste da Razão de verossimilhança (likelihood-ratio test Pawitan (2001), pág. 258), consiste em designar uma região crítica baseada na estatística

$$\Lambda(x) = \frac{\sup\{L(x|\theta) : \theta \in \Theta_{H_0}\}}{\sup\{L(x|\theta) : \theta \in \Theta\}}, \tag{5.18}$$

onde L é a função de verossimilhança, x representa o conjunto de dados observados e Θ e Θ_{H_0} representam o espaço paramétrico e o subconjunto de Θ correspondente à hipótese H_0, respectivamente. Como $\Theta_{H_0} \subset \Theta$, a estatística $\Lambda(x)$ assume valores no intervalo $[0, 1]$. Quanto maior esse valor, maior a evidência a favor da hipótese. A aplicabilidade desse teste (em geral) se deve ao Teorema de Wilks Wilks (1938), o qual afirma que a distribuição de $-2\log\Lambda(x)$ (sob a hipótese H_0) aproxima-se assintoticamente à distribuição χ^2 (qui-quadrado) com grau de liberdade igual à diferença entre as dimensões de Θ e Θ_{H_0}. Isso possibilita a obtenção de aproximações do p-valor por meio da substituição da distribuição de $\Lambda(x)$ pela distribuição qui-quadrado.

5.4 FBST vs LR

As inconsistências geradas pela necessidade de incluir hipóteses precisas no contexto da teoria bayesiana ortodoxa (Seção 5.1) induzem ao uso do FBST em lugar dos demais testes apresentados acima. Em contrapartida, o BIC e o LR são de fácil implementação, o que não ocorre com o FBST. Isso, obviamente, se deve à indisponibilidade de algoritmos eficientes para o cálculo de integrais numéricas em dimensões elevadas[5]. Para esse fim, algoritmos baseados em métodos de Monte Carlo (Robert e Casella (2004)) parecem ser os mais adequados. A seguir, comparamos as performances do FBST e do LR para um teste de hipótese específico.

Descrição do experimento.

O teste consiste em testar a plausibilidade do modelo de Dirichlet linear (MDL) em relação ao modelo de Dirichlet quadrático (MDQ) da Seção 4.1, isto é:

$$H_0 : \beta_{i3} = 0, i = 1, 2, 3 \quad vs \quad H : \beta_{ij} = 1, i, j = 1, 2, 3. \tag{5.19}$$

Para cada valor de $n \in \{30, 40, 50, 60\}$, foram geradas dois grupos $\mathcal{C}_1$ e $\mathcal{C}_2$ com $m = 500$ amostras independentes com n elementos cada(com reposição). As amostras do grupo $\mathcal{C}_1$ e $\mathcal{C}_2$ correspondem, respectivamente, a conjuntos de dados composicionais obtidos pelas fórmulas (2.2) e (2.6) (mais uma pequena perturbação normal com média zero) utilizando os parâmetros de máxima verossimilhança correpondentes aos modelos (MDL) e (MDQ) da tabela 4.7.

O desempenho de cada teste (FBST e LR) pode ser medido analizando-se as performances dos erros dos Tipos I e II (rejeitar a hipótese H_0, sendo ela verdadeira e não rejeitar a hipótese H_0, sendo ela falsa, respectivamente.). Para cada amostra em $\mathcal{C}_1$ e em $\mathcal{C}_2$, a hipótese H_0 foi aceita ou rejeitada com base nas aproximações assintóticas para o p-valor descritas nas sessões anteriores[6],

[5] Observe, por exemplo, que o modelo de Dirichlet linear empregado para modelar as porcentagens de casamentos por faixa etária apresenta 42 parâmetros!

[6] Para o cálculo do e-valor, utilizamos uma priori uniforme imprópria em $\mathbb{R}^{C\times D}$ e um algoritmo de integração

para o erro do Tipo I limitado a $\alpha = 5\%$. Como as amostras provenientes de $\mathcal{C}_1$ e de $\mathcal{C}_2$ foram geradas dentro e fora de H_0, respectivamente, a porcentagem das amostras de $\mathcal{C}_1$ (respectivamente de $\mathcal{C}_2$) para as quais H_0 foi rejeitada (respectivamente aceita) fornece uma estimativa para o erro do Tipo I (respectivamente Tipo II).

Uma outra abordagem consiste em analisar o comportamento do erro do Tipo II produzido (em termos da taxa de aceitação das amostras em $\mathcal{C}_2$) ao selecionar os pontos de corte que determinam as regiões críticas de rejeição de tal forma que, com esses limiares estabelecidos, a hipótese H_0 seja rejeitada em apenas 5% das amostras em $\mathcal{C}_1$.

Os resultados seguem abaixo.

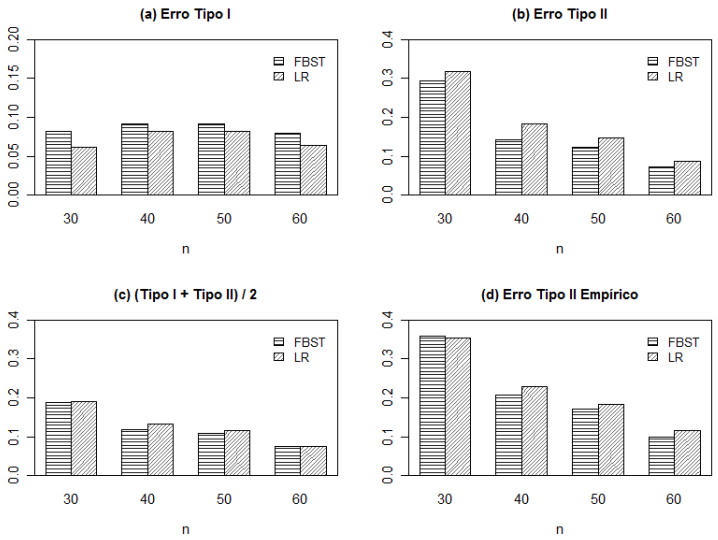


Figura 5.1: *(a) Erro Tipo I, (b) Erro Tipo II, (c) Erro médio e (d) Erro Tipo II empírico.*

Embora o FBST tenha apresentado uma taxa de erro do Tipo I ligeiramnete maior do que o LR, os erros do Tipo II (assintótico e empírico) e erro médio referentes ao FBST apresentaram valores infeirores aos respectivos referentes ao LR, sugerindo um poder discriminatório maior para o FBST.

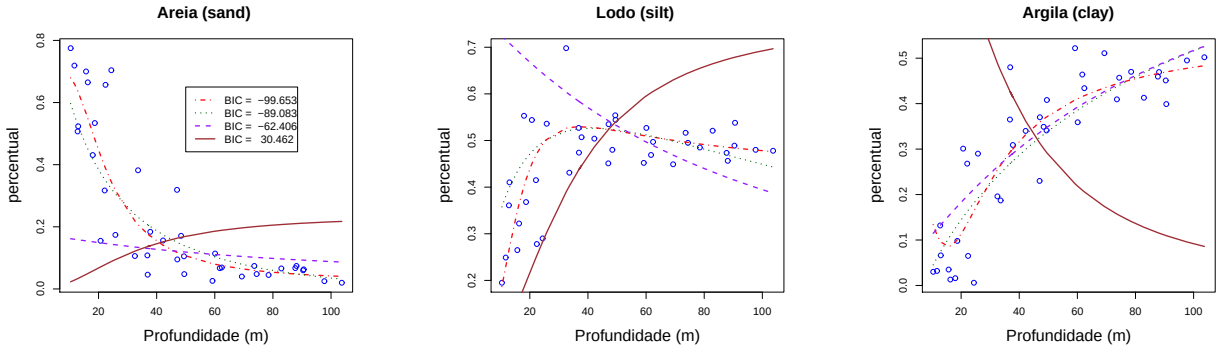
5.5 Aplicações do BIC aos modelos de Dirichlet linear

5.5.1 Lago Ártico

A Figura 5.2 mostra as previsões geradas pelo melhor (menor BIC), pior (maior BIC) e alguns outros modelos correspondentes a valores intermediários de BIC dentre todos os submodelos[7] do modelo de Dirichlet polinomial quadrático MDQ da Seção 4.1 (os valores correspondentes de BIC para os modelos MDL e MDQ são -95.726 e -82.994).

numérica do tipo Metropolis-Hasting com distribuição normal multivariada como distribuição proposta Martin *et al.* (2011).

[7]Entende-se por submodelo de um modelo *mod* a todo modelo obtido a partir de *mod* zerando-se, eventualmente, algumas das entradas da sua matriz de modelo.



(a) (b) (c)

Figura 5.2: *Submodelos de MDQ*

Observa-se que, conforme o valor do BIC decai, o ajuste é comprometido. Esse fato reitera que os modelos que produzem ajustes ruins possuem pouca chance de ser o modelo correto (são pouco prováveis *à posteriori*, de acordo com Schwarz).

Matriz de modelo	Estimativas	BIC	Posição
$\begin{matrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{matrix}$	$\begin{matrix} 3.545e+0 & - & 4.452e-4 \\ - & - & 9.134e-3 \\ 3.033e+0 & -3.524e+1 & 1.239e+2 \end{matrix}$	-99.653	1
$\begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{matrix}$	$\begin{matrix} 5.168e+0 & -2.986e-2 & - \\ - & 2.800e-1 & - \\ -5.264e-1 & 5.831e-2 & 2.687e-3 \end{matrix}$	-89.083	100
$\begin{matrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{matrix}$	$\begin{matrix} - & 3.682e-2 & - \\ - & 1.648e-1 & - \\ 3.887e-2 & - & 2.157e-3 \end{matrix}$	-62.406	200
$\begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{matrix}$	$\begin{matrix} - & - & 1.830e-4 \\ - & - & 5.865e-4 \\ 7.771e-1 & - & - \end{matrix}$	30.46160	343

Tabela 5.1: *Alguns submodelos de MDQ ordenados do menor para o maior BIC*

5.5.2 Despesas domésticas

Intuitivamente, é de se esperar que homens e mulheres apresentem comportamentos distintos com relação aos gastos com as despesas domésticas. No exemplo 4.2, essa suspeita é incorporada ao modelo de Dirichlet linear por meio da função $f_2(r,s)$. Dessa forma, a dependência (ou independência) da distribuição dos gastos com relação ao sexo pode ser matematicamente formulada pela hipótese:

$$H_0 : \beta_{i2} = 0, i = 1,2,3,4 \quad vs \quad H : \beta_{i2} \neq 0, i = 1,2,3,4. \tag{5.20}$$

Os BICs relativos aos modelos descritos por H_0 e por H são iguais a -100.457 e -172.187, respectivamente, donde conclui-se que a hipótese H_0 deve ser rejeita. Para esse exemplo, em especial, não

seria necessário tanta sofisticação, pois essa conclusão é imediata a partir de uma análise gráfica dos dados observados. Não há, porém, apelos visuais para determinar (no caso geral) quais das funções $f_1(s,r) = 1$, $f_2(s,r) = s$, $f_3(s,r) = r$, $f_4(s,r) = r^2, f_2(s,r) = \log(r)$ são relevantes para esse modelo. Segundo o critério de Schwarz, o melhor modelo (BIC = -178.779) corresponde ao uso das funções f_1, f_2 e f_4, apenas.

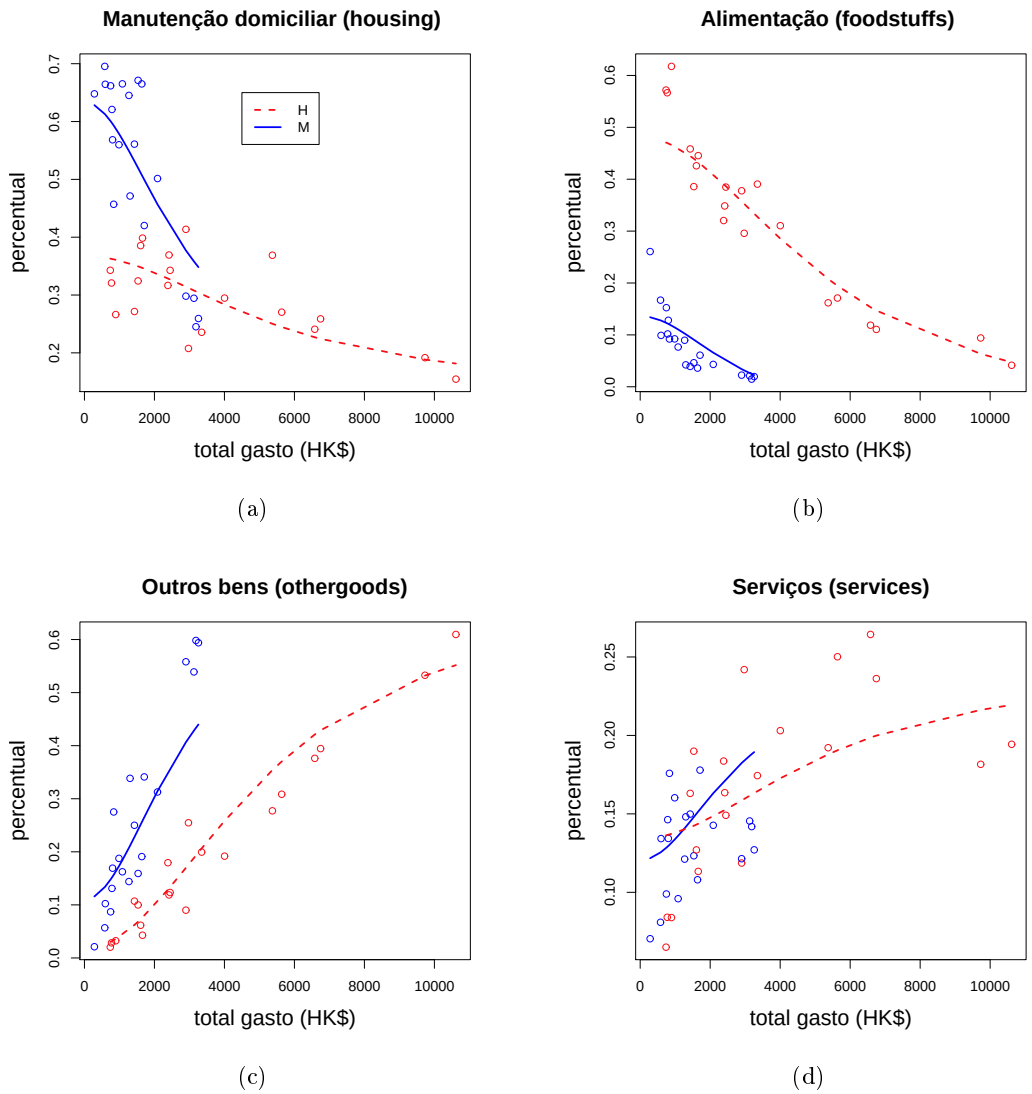


Figura 5.3: *Melhor modelo segundo Schwarz*

Capítulo 6

Conclusões

Neste trabalho realizamos um estudo sistemático dos modelos de regressão sobre dados composicionais descritos nos trabalhos de Hijazi e Jernigan (2009), Melo *et al.* (2009), Aitchinson (1986) e Wang *et al.* (2007). O primeiro assume que a distribuição condicional da variável resposta, dado o valor da variável explicativa, obedece um Modelo de Dirichlet cujos parâmetros dependem linearmente das variáveis explicativas (covariáveis). O segundo é uma variante do primeiro, no qual funções positivas são aplicadas às combinações lineares das covariáveis para a obtenção dos parâmetros. O terceiro assume que a distribuição condicional dos dados transformados (via transformação Logaritmo da Razão), dado o valor da variável explicativa, segue um modelo Normal multivariado cuja média depende do valor da variável explicativa. O último parametriza o Simplex utilizando coordenadas esféricas generalizadas e ajusta os ângulos diretamente utilizando um modelo de regressão linear.

Também abordamos três métodos de seleção de modelos: BIC, FBST e Teste da razão de verossimilhança.

As principais contribuições obtidas nesse estudo são:

1. A elaboração de um algoritmo mais eficiente para a obtenção de um ponto inicial para o processo iterativo de maximização da função de verossimilhança no Modelo de Dirichlet Linear (Seção 2.3).

2. A completa caracterização das transformações equivalentes (Seção 3.1.1) e a elaboração de novas transformações entre o Simplex e o Espaço eulidiano (Seção 3.3).

3. Um estudo detalhado do critério de Schwarz (BIC).

Os exemplos apresentados no Capítulo 4 mostram que os modelos de Dirichlet Linear, de Melo, Vasconcellos e Lemonte e os modelos baseados em transformações entre o Simplex e o Espaço Euclidiano possuem grande versatilidade. Em alguns casos, porém, a qualidade do ajuste foi devida à escolha artificial da forma de $x_\bullet$. Não parece natural, por exemplo, introduzir a componente $m * cos(t/100)$ no modelo de Dirichlet linear para o exemplo dos Casamentos por faixa etária. Embora os métodos de seleção de modelo apresentados no Capítulo 5 sirvam para discernir quais das componentes de $x_\bullet$ são relevantes, o desafio de escolher adequadamente a forma de $x_\bullet$, para cada aplicação, ainda permanece.

Nesse sentido, os modelos baseados em transformações entre o Simplex e o Espaço euclidiano poderiam ser considerados ligeiramente superiores, pois conseguimos bons ajustes utilizando so-

mente funções polinomiais e logaritmicas. Além disso, tais modelos destacam-se sobre os demais pela fácil implementação e por produzir estimativas sempre dentro do Simplex. Os modelos esféricos, em particular, poderiam ser considerados ligeiramnete superiores ao modelo Logaritmo da Razão de Aitchison por serem sensíveis a permutações das componentes das composições (ver Seção 3.3.1).

O modelo de Melo, Vasconcellos e Lemonte também produz estimativas sempre dentro do Simplex. Porém, além do uso de funções artificiais, a escolha natural para as funções positivas g_j, a função exponencial, gera grande instabilidade numérica. Uma alternativa é o uso da função arcotangente, como já mencionado na Seção 2.4.

Os modelos de Dirichlet Linear e de Wang podem ser considerados ligeiramente inferiores por não produzirem estimativas sempre dentro do Simplex. O caso do modelo de Dirichlet Linear, entretanto, não é tão crítico como o do modelo de Wang (ver Seção 3.2). Por outro lado, o Algoritmo (2.3.2) garante uma maior estabilidade ao modelo de Dirichlet Linear, destacando-se sobre o modelo de Melo, Vasconcellos e Lemonte.

Sobre os métodos de seleção de modelos, podemos dizer que não há uma relação de preferência evidente.

Embora a comparação das performances do FBST e do LR induza a uma pequena preferência pelo FBST, podemos considerar que o contraste entre as complexidades de implementação anula um possível favoritismo. Consideramos, por outro lado, que o BIC e o FBST não são concorrentes, mas sim métodos complementares. Pois o BIC não parece ser designado para comparar modelos encaixados (a especialidade do FBST).

Esperamos, num futuro próximo, disponibilizar as rotinas desenvolvidas para a construção dos modelos apresentados como pacotes do software livre de Estatística computacional R.

Apêndice A

Cálculo dos coeficientes da regressão linear por mínimos quadrados

Teorema 2. *Se a matriz X de* (1.1) *possui posto máximo, então o residuo quadrático* (3.5) *possui um único ponto de mínimo $\widehat{\beta}$. Se, além disso, a matriz de modelo* (2.3) *é completa e $(\widetilde{\beta}, \widetilde{\Sigma})$ é a estimativa de máxima verossimilhança de* (3.4) *(supondo que exista e seja única), então $\widetilde{\beta} = \widehat{\beta}$.*

Prova:

Fazendo $M_{x_\bullet} = \begin{bmatrix} x_\bullet & 0 & \dots & 0 \\ 0 & x_\bullet & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_\bullet \end{bmatrix}$, temos que

$$\mu(x_\bullet, \beta) = psi(x_\bullet, \beta), M_{x_\bullet} \times \beta_v, \tag{A.1}$$

para $\beta_v = \Big(\ \beta_{11}, \beta_{12}, \dots, \beta_{1k}, \beta_{21}, \beta_{22}, \dots, \beta_{2k}, \dots, \beta_{d1}, \beta_{d2}, \dots, \beta_{dk} \ \Big)$.

Logo, (3.5) assume a forma:

$$Res^2(\beta) \ = \left\| \underbrace{\begin{bmatrix} y_{1\bullet}^{(tr)} \\ y_{2\bullet}^{(tr)} \\ \dots \\ y_{n\bullet}^{(tr)} \end{bmatrix}}_{Y_v} - \underbrace{\begin{bmatrix} M_{x_{1\bullet}} \\ M_{x_{2\bullet}} \\ \vdots \\ M_{x_{n\bullet}} \end{bmatrix}}_{M} \times \beta_v \right\|^2 \tag{A.2}$$

Reescrevendo (A.2) na forma matricial, obtemos $Res^2(\beta) = [Y_v - M \times \beta_v]^t \times [Y_v - M \times \beta_v] =$

$$\beta_v^t \times \underbrace{[M^t \times M]}_{A} \times \beta_v - 2 \underbrace{[Y_v^t \times M]}_{b} \times \beta_v + \underbrace{Y_v^t \times Y_v}_{c} \tag{A.3}$$

Como X possui posto máximo, então[1] A é positiva definida, e, portanto, a forma quadrática (A.3) assume um único mínimo (no ponto [2] $\widehat{\beta_v} = A^{-1} \times b^t$).

Por (3.4), temos que

$$logL(\beta, \Sigma) = n * log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right) - \frac{1}{2}\sum_{i=1}^{n}(T(y_{i\bullet}) - \mu(x_{i\bullet}))^t \times \Sigma^{-1} \times (T(y_{i\bullet}) - \mu(x_{i\bullet})) \quad \text{(A.4)}$$

e, portanto,

$$\frac{\partial \log L}{\partial \beta_{ij}} = -\sum_{l=1}^{n}\left[\begin{array}{ccccc} 0 & \dots & \underbrace{x_{lj}}_{\text{posição i}} & \dots & 0 \end{array}\right] \times \Sigma^{-1} \times (T(y_{l\bullet}) - \mu(x_{l\bullet})) \ .$$

Temos, então, que o sistema $\frac{\partial \log L}{\partial \beta_{ij}}(\beta,\ \Sigma) = 0,\ i = 1, 2, \dots, d$ (j fixado) equivale a que $\Sigma^{-1} \times \left(\sum_{l=1}^{n} x_{lj}(y_{l\bullet}^{(tr)} - \mu(x_{l\bullet}))\right)$ seja ortogonal[3] a $\mathbb{R}^d$, donde segue que

$$\sum_{l=1}^{n} x_{lj}(y_{l\bullet}^{(tr)} - \mu(x_{l\bullet})) = 0. \quad \text{(A.5)}$$

Como $logL(\beta, I_d) = n * log\left(\frac{1}{(2\pi)^{d/2}}\right) - Res^2(\beta)$, então $\widehat{\beta}$ satisfaz (A.5) ($\forall\ j$) e, como (A.5) não depende de Σ, segue que

$$\frac{\partial \log L}{\partial \beta_{ij}}\left(\widehat{\beta},\ \widetilde{\Sigma}\right) = 0\ \forall\ i, j \quad \text{(A.6)}$$

Pela unicidade da estimativa de máxima verossimilhança, obtemos $\widehat{\beta} = \widetilde{\beta}$.

□

[1] Note que M tem posto máximo, pois, rearranjando-se convenientemente as linhas de M, obtemos $\begin{bmatrix} X & & & \\ & X & & \\ & & \ddots & \\ & & & X \end{bmatrix}$, que tem posto máximo. Logo, a trasformação $z \longmapsto M \times z$ é injetora e, portanto, $z^t \times A \times z = ||M \times z||^2 > 0$, for $z \neq 0$.

[2] Uma simples manipulação algébrica mostra que $Res(\widehat{\beta} + \epsilon) = Res(\widehat{\beta}) + \epsilon^t \times A \times \epsilon$ (Watkins (2002), pág. 559).

[3] Evidentemente, isso só vale se a matriz de modelo (2.3) for completa.

Apêndice B

Base de dados

A seguir apresentamos os conjuntos de dados utilizados nesta dissertação.

1. O primeiro é referente à 39 composições do solo do Lago Ártico (areia, lodo e argila em função da profundidade) apresentados por Coakley e Rust Coakley e Rust (1968) e adaptados por Aitchinson (1986).

2. O segundo é referente ao conjunto de dados encontrado em Aitchinson (1986), constituído pelas despesas domésticas (separadas em Manutenção domiciliar, Alimentação, Outros bens e Serviços.) de 20 homens e 20 mulheres, em função do total gasto.

3. O terceiro corresponde a um conjunto de dados fornecido pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e disponível em *http://seriesestatisticas.ibge.gov.br/* relativo à variação das porcentagens de casamentos por faixa etária e sexo no período 1984-2002.

B.1 Lago Ártico

Profundidade(m)	Areia	Lodo	Argila	Profundidade(m)	Areia	Lodo	Argila
10.40	0.78	0.20	0.03	47.10	0.10	0.54	0.37
11.70	0.72	0.25	0.03	48.40	0.17	0.48	0.35
12.80	0.51	0.36	0.13	49.40	0.10	0.55	0.34
13.00	0.52	0.41	0.07	49.50	0.05	0.54	0.41
15.70	0.70	0.26	0.04	59.20	0.03	0.45	0.52
16.30	0.66	0.32	0.01	60.10	0.11	0.53	0.36
18.00	0.43	0.55	0.02	61.70	0.07	0.47	0.46
18.70	0.53	0.37	0.10	62.40	0.07	0.50	0.43
20.70	0.15	0.54	0.30	69.30	0.04	0.45	0.51
22.10	0.32	0.41	0.27	73.60	0.07	0.52	0.41
22.40	0.66	0.28	0.06	74.40	0.05	0.49	0.46
24.40	0.70	0.29	0.01	78.50	0.04	0.48	0.47
25.80	0.17	0.54	0.29	82.90	0.07	0.52	0.41
32.50	0.11	0.70	0.20	87.70	0.07	0.47	0.46
33.60	0.38	0.43	0.19	88.10	0.07	0.46	0.47
36.80	0.11	0.53	0.36	90.40	0.06	0.49	0.45
37.80	0.18	0.51	0.31	90.60	0.06	0.54	0.40
36.90	0.05	0.47	0.48	97.70	0.02	0.48	0.49
42.20	0.16	0.50	0.34	103.70	0.02	0.48	0.50
47.00	0.32	0.45	0.23				

Tabela B.1: *Composição do solo do Lago Ártico em função da profundidade.*

B.2 Despesas Domésticas

	T	M	A	O	S		T	M	A	O	S
Homens	738	0.34	0.57	0.02	0.07	Mulheres	284	0.65	0.26	0.02	0.07
	773	0.32	0.57	0.03	0.08		581	0.70	0.17	0.06	0.08
	894	0.27	0.62	0.03	0.08		596	0.66	0.10	0.10	0.13
	1429	0.27	0.46	0.11	0.16		748	0.66	0.15	0.09	0.10
	1532	0.32	0.39	0.10	0.19		786	0.62	0.10	0.13	0.15
	1606	0.39	0.43	0.06	0.13		804	0.57	0.13	0.17	0.13
	1659	0.40	0.45	0.04	0.11		836	0.46	0.09	0.28	0.18
	2385	0.32	0.32	0.18	0.18		986	0.56	0.09	0.19	0.16
	2416	0.37	0.35	0.12	0.16		1084	0.67	0.08	0.16	0.10
	2448	0.34	0.38	0.12	0.15		1271	0.65	0.09	0.14	0.12
	2899	0.41	0.38	0.09	0.12		1303	0.47	0.04	0.34	0.15
	2972	0.21	0.30	0.25	0.24		1428	0.56	0.04	0.25	0.15
	3349	0.24	0.39	0.20	0.17		1533	0.67	0.05	0.16	0.12
	4004	0.29	0.31	0.19	0.20		1639	0.67	0.04	0.19	0.11
	5371	0.37	0.16	0.28	0.19		1709	0.42	0.06	0.34	0.18
	5637	0.27	0.17	0.31	0.25		2088	0.50	0.04	0.31	0.14
	6582	0.24	0.12	0.38	0.26		2899	0.30	0.02	0.56	0.12
	6748	0.26	0.11	0.39	0.24		3128	0.29	0.02	0.54	0.15
	9731	0.19	0.09	0.53	0.18		3186	0.25	0.01	0.60	0.14
	10615	0.15	0.04	0.61	0.19		3258	0.26	0.02	0.59	0.13

Tabela B.2: *Despesas domésticas: T = Total gasto (em HK$); A = Alimentação; O = Outros Bens; S = Serviços.*

B.3 Casamentos por faixa-etária

	Ano	0 a 19	20 a 29	30 a 39	40 a 49	50 a 59	60+.
Homens	1984	7.41	69.12	15.34	4.56	2.17	1.40
	1985	7.53	70.25	14.58	4.15	2.07	1.41
	1986	7.90	70.09	14.25	4.09	2.15	1.51
	1987	8.12	71.44	13.87	3.44	1.77	1.36
	1988	7.65	70.18	14.78	3.95	1.98	1.44
	1989	7.96	73.08	13.59	2.73	1.39	1.24
	1990	7.82	72.46	14.23	2.78	1.42	1.29
	1991	7.69	71.69	14.97	2.88	1.41	1.35
	1992	7.47	70.65	15.77	3.15	1.52	1.44
	1993	6.99	70.01	16.63	3.24	1.55	1.59
	1994	5.10	68.70	19.20	3.61	1.63	1.78
	1995	6.89	67.60	18.32	3.72	1.69	1.77
	1996	6.63	66.43	19.14	4.10	1.79	1.91
	1997	6.23	65.66	19.98	4.31	1.84	1.96
	1998	5.85	64.83	20.67	4.58	1.97	2.10
	1999	5.43	62.88	21.80	5.37	2.28	2.24
	2000	5.27	62.27	22.24	5.56	2.33	2.31
	2001	4.17	60.36	23.97	6.25	2.65	2.60
	2002	3.81	59.34	24.54	6.73	2.81	2.78
Mulheres	1984	33.36	51.70	10.07	3.20	1.23	0.44
	1985	33.76	52.29	9.45	2.94	1.15	0.42
	1986	33.92	52.26	9.19	3.00	1.18	0.45
	1987	34.27	53.20	8.62	2.53	0.97	0.41
	1988	32.72	53.34	9.54	2.86	1.09	0.43
	1989	34.71	54.34	7.89	1.94	0.75	0.36
	1990	33.94	54.66	8.29	1.96	0.78	0.37
	1991	33.39	54.51	8.90	2.03	0.77	0.41
	1992	32.33	54.74	9.48	2.23	0.81	0.41
	1993	31.38	55.03	9.93	2.30	0.89	0.47
	1994	31.20	54.29	10.58	2.47	0.96	0.50
	1995	30.86	53.80	11.05	2.65	1.09	0.55
	1996	30.16	53.47	11.67	2.94	1.16	0.60
	1997	28.96	53.89	12.19	3.06	1.17	0.74
	1998	28.01	53.86	12.78	3.32	1.29	0.74
	1999	26.13	53.73	14.03	3.89	1.51	0.72
	2000	25.02	54.10	14.57	4.03	1.54	0.74
	2001	21.27	55.38	16.10	4.64	1.79	0.83
	2002	19.64	55.70	16.86	4.99	1.93	0.88

Tabela B.3: *Porcentagen de casamentos por faixa etária e sexo. Fonte: IBGE*

Referências Bibliográficas

Aitchinson(1986) J. Aitchinson. *The Statistical Analysis of Compositional Data.* Chapman and Hall. Citado na pág. 2, 13, 15, 51, 55

Aitchison(1982) J. Aitchison. The statistical analysis of ompositiona data. *Journal of the Royal Statistical Society*, 44:139–177. Citado na pág. 13, 14

Aitchison e Shen(1980) J. Aitchison e S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67:261–272. Citado na pág. 13

Aitchison *et al.***(2000)** J. Aitchison, C. Barceló-Vidal, J. A. Martín-Fernández, e V. Pawlowsky-Glahn. Logratio analysis and compositional distance. *Mathemathical Geology*, 32:271–275. Citado na pág. 15, 24

Campbell e Mosimann(1987) G. Campbell e J. Mosimann. Multivariate methods for proportional shape. Em *ASA Proceedings of the Section on Statistical Graphics*, páginas 10–17. Citado na pág. 2, 6

Coakley e Rust(1968) J. P. Coakley e B. R. Rust. Sedimentation in a arctic lake. *Sedimentary Petrology*, 38:1290–1300. Citado na pág. 2, 55

DeGroot e Schervish(2002) M. H. DeGroot e M. J. Schervish. *Probability and Statistics.* Addison - Wesley, terceira edição. Citado na pág. 45

Draper e Smith(1998) N. R. Draper e H. Smith. *Applied regression analysis.* John Willwy & Sons, Inc., terceira edição. Citado na pág. 39

Evans(1997) M. Evans. Bayesian inference procedures via the concept of relative surprise. *Communications in Statistics*, 26:11251143. Citado na pág. 45

Hijazi e Jernigan(2009) H. H. Hijazi e R. W. Jernigan. Modeling compositional data using dirichlet regression models. *Journal of Applied Probability & Statistics*, 4:77–91. Citado na pág. 7, 25, 51

Krätschmer(2007) V. Krätschmer. The uniqueness of extremum estimation. *Statistics & Probability Letters*, 77:942–951. Citado na pág. 11

Madruga *et al.***(2003)** M.R. Madruga, C. A. B. Pereira, e J. M. Stern. Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference*, 117:185–198. Citado na pág. 45

Martin *et al.***(2011)** A. D. Martin, K. M. Quinn, e J. H. Park. Mcmcpack: Markov chain monte carlo in r. *Journal of Statistical Software: http://www.jstatsoft.org/v42/i09/*, 42:1–21. Citado na pág. 47

Melo *et al.***(2009)** T. F.N. Melo, K. L.P. Vasconcellos, e A. J. Lemonte. Some restriction tests in a new class of regression models for proportions. *Computational Statistics and Data Analysis*, 53: 3972–3979. Citado na pág. 2, 10, 39, 51

Pawitan(2001) Y. Pawitan. *In all Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford University Press. Citado na pág. 43, 46

Pereira e Stern(2008) C. A. B. Pereira e J. M. Stern. Special characterizations of standart discrete models. *Revstat*, 6:199–230. Citado na pág. 2, 5, 7, 8

Pereira e Stern(1999) C. A. B. Pereira e J. M. Stern. Evidence and credibilidty: Full bayesian significance test of precise hypothesis. *Entropy Journal*, 1:99–110. Citado na pág. 45

Pereira *et al.***(2008)** C. A. B. Pereira, J. M. Stern, e S. Wechsler. Can a significant test be genuinely bayesian. *Bayesian Analysis*, 3:79–100. Citado na pág. 45

Robert e Casella(2004) C. P. Robert e G. Casella. *Monte Carlo Statistical Methods.* Springer, segunda edição. Citado na pág. 46

Ronning(1989) G. Ronning. Maximum likelihood estimation of dirichlet distributions. *Journal of Statistical Computation and Simulation*, 32:215–221. Citado na pág. 5

Schervish(1997) M. J. Schervish. *Theory of Statistics.* Springer. Citado na pág. 45

Schwarz(1978) G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464. Citado na pág. 39, 43

Stern(2008) J. M. Stern. *Tutorial: Cognitive Construtivism and the Epistemic Significance of Sharp Statistical Hypotheses.* 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engeneering, Boracéia-Brazil. Citado na pág. 41

Wang *et al.***(2007)** H. Wang, Q. Liu, H. M.K. Mok, L. Fu, e W. M. Tse. A hyperspherical transformation forecasting model for compositional data. *European Journal of Operational Research*, 179:459–468. Citado na pág. 2, 18, 51

Watkins(2002) D. S. Watkins. *Fundamentals of Matrix Computations.* John Wiley & Sons. Citado na pág. 54

Wilks(1938) S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9:60–62. Citado na pág. 46

Índice Remissivo

dado composicional, 1
distribuição
    de Dirichlet, 5, 6
    normal, 13

Familia Exponencial, 5
função
    de verossimilhança, 7

isomorfismo linear, 15

matriz
    de modelo, 6

norma
    euclidiana, 8

processos de Bernoulli, 9

regressão, 1
regressão
    de Dirichlet, 5
    linear, 6

Simplex, 1

transformação
    Log-esférica, 22
    Logaritmo da razão, 13
    Tangente-esférica, 22
transformações
    equivalentes, 15
    logísticas, 15

verossimilhança, 8