

# Resumo

O principal objectivo desta tese é a construção de novos procedimentos robustos para a análise de *componentes principais comuns* (CPC). Este modelo, proposto por Flury (1984), é uma generalização das *componentes principais* (CP) para  $k$  grupos e assume que as  $k$  matrizes de covariâncias têm valores próprios diferentes mas vectores próprios idênticos. A alternativa robusta mais intuitiva para a estimação é a técnica de “plug-in” (PI). Outra alternativa robusta é a baseada em técnicas de “projection-pursuit” (PP). Nesta tese é avaliado o grau de robustez destas alternativas.

Nos Capítulos 2 e 3 apresentamos resultados de vários autores que consideramos importantes para a exposição dos capítulos seguintes. A robustez dos dois procedimentos (PI e PP) é avaliada no Capítulo 4 através das *funções de influência parciais* dos funcionais valores e vectores próprios comuns. Estas funções são uma generalização da *função de influência* das CP. A partir destas expressões também obtivemos as variâncias assintóticas dos estimadores PI e PP. No Capítulo 5 é proposto um procedimento robusto PP geral através da aplicação duma função “score”  $f(t)$ . São derivadas as *funções de influência parciais* destes funcionais e deduzidas as variâncias assintóticas. Foi possível demonstrar, para o caso particular do modelo proporcional, que os estimadores PP gerais atingem o valor mínimo da variância assintótica quando  $f(t) = \ln(t)$ . Também no Capítulo 5, o desempenho dos novos estimadores PP é comparado com os usuais com base em simulações de Monte Carlo. A eficácia destes estimadores PP é ainda avaliada em conjuntos de dados reais. No Capítulo 6 são propostas duas novas medidas para a detecção de observações discordantes/influentes na estimação das CP e das CPC, construídas a partir das *funções de influência parciais* dos valores e vectores próprios comuns. A sua eficácia foi avaliada em dois conjuntos de dados reais. Finalmente no Capítulo 7 apresentamos as conclusões do trabalho e identificamos alguns tópicos para eventuais trabalhos futuros.

**Palavras Chave:** *Componentes principais comuns*, Estimação robusta, Estimadores de “plug-in”, Estimadores de “projection-pursuit” gerais, *Funções de influência parciais*, Observações discordantes/influentes.



# Abstract

In this thesis we propose robust procedures for the *common principal components* (CPC) model. This model, proposed by Flury (1984), is a generalization of *principal components* (CP) to  $k$  populations and assumes that the  $k$  covariance matrices have different eigenvalues but identical eigenvectors. The robust plug-in (PI) technique is perhaps the most simple and intuitive robust approach. Another robust approach is based on projection-pursuit (PP) principles. In this thesis we evaluate the robustness of both procedures.

Chapters 2 and 3 survey results from several authors necessary for the remaining chapters. With the aim of evaluating the robustness of both approaches (PI and PP) in Chapter 4 we derive the *partial influence functions* for the eigenvalues and the eigenvectors functionals. These expressions then allow the heuristic derivation of the asymptotic variances for the PI and the PP estimators. In Chapter 5 a general PP approach obtained with the application of a score function  $f(t)$  is proposed. We derive the *partial influence functions* for this functionals and the corresponding asymptotic variances. It was also possible to prove that, for the particular case of a proportional model, the general PP estimators with  $f(t) = \ln(t)$  are optimal in the sense of minimizing the asymptotic variance. Through a simulation study these last PP estimators are compared with those related to  $f(t) = t$ . Chapter 5 also presents the performance of the general PP estimators for real data sets. Chapter 6 proposes two detection measures of outlying/influential observations for the CPC and the CP models, based on the *partial influence functions* for the eigenvalues and the eigenvectors functionals. The performance of the two new measures is evaluated in real data sets. Finally, Chapter 7 states the main conclusions and discusses future work.

**Key words:** *Common principal components*, General projection-pursuit estimators, *Partial influence functions*, Plug-in estimators, Outlying/influential observations, Robust estimation.



# Agradecimentos

Foram várias as pessoas que me ajudaram a terminar com este projecto. No entanto, gostaria de prestar um agradecimento especial

À minha orientadora, Professora Ana Pires, pelo seu constante apoio, o estímulo e orientação que prestou ao longo deste percurso académico. Além disso, e não menos importante, não posso deixar de referir a sua “infinita” paciência e o enorme afecto demonstrado.

À Professora Graciela Boente pela cedência dos programas MATLAB, pelas sugestões e visitas que efectuou ao Departamento de Matemática do Instituto Superior Técnico e Centro de Matemática e Aplicações (CEMAT) e o trabalho conjunto desenvolvido.

Ao Professor Jorge Buescu pela sugestão do Princípio do Máximo e das função harmónicas.

Ao Engenheiro M. J. Pimentel Pereira por autorizar a utilização dos dados das Variedades de Castanheiros.

Aos meus colegas da Secção de Estatística e Aplicações Cláudia Nunes, Conceição Amado, Delfina Barbosa, Manuel Morais e Rosário Oliveira pelo carinho, o incentivo e a compreensão demonstrada.

À minha família que me “aturou” vários dias de mau humor. Não posso também deixar de agradecer a ajuda do Vitor e do seu pai.

Ao CEMAT pelo apoio financeiro oferecido que permitiu quer a participação em conferências como a compra de livros e computadores, tudo isto imprescindível para a conclusão desta investigação.

Ao Fundo Social Europeu que através da Acção 5.3 - Formação Avançada de Docentes do Ensino Superior - do Programa PRODEP III co-financiou esta tese de doutoramento.

A todos muito obrigada.



# Índice

<b>Resumo</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Agradecimentos</b>	<b>v</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Figuras</b>	<b>xv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objectivo desta tese . . . . .	1
1.2 Metodologia utilizada . . . . .	5
<b>2 Estruturas de Covariâncias</b>	<b>9</b>
2.1 Introdução . . . . .	9
2.2 Algumas estruturas de covariâncias . . . . .	10
2.3 As componentes principais comuns . . . . .	14
2.3.1 Estimação dos parâmetros por máxima verosimilhança . . . . .	17
2.3.2 Propriedades dos estimadores . . . . .	19
2.3.3 Exemplo . . . . .	21
2.3.4 Revisão bibliográfica . . . . .	24
<b>3 Estimação Robusta</b>	<b>41</b>

3.1	Introdução . . . . .	41
3.2	Algumas generalidades . . . . .	42
3.3	Observações discordantes e erros grosseiros . . . . .	47
3.4	A abordagem minimax de Huber . . . . .	49
3.5	A abordagem infinitesimal de Hampel . . . . .	51
3.5.1	Robustez qualitativa . . . . .	52
3.5.2	Função de influência . . . . .	53
3.5.3	Medidas de robustez derivadas da função de influência . . . . .	58
3.5.4	Versões empíricas da função de influência . . . . .	60
3.5.5	Ponto de ruptura . . . . .	61
3.6	Estimadores de dispersão univariados . . . . .	63
3.7	Algumas generalidades sobre estimação multivariada . . . . .	67
3.8	Estimadores de dispersão multivariados . . . . .	69
<b>4</b>	<b>Estimação Robusta das Componentes Principais Comuns</b>	<b>79</b>
4.1	Introdução . . . . .	79
4.2	Revisão bibliográfica . . . . .	80
4.3	Novos resultados . . . . .	89
4.3.1	Funções de influência parciais e variâncias assintóticas dos estimadores “plug-in” . . . . .	96
4.3.2	Funções de influência parciais e variâncias assintóticas dos estimadores “projection–pursuit” . . . . .	108
<b>5</b>	<b>Estimadores “Projection–Pursuit” Gerais</b>	<b>125</b>
5.1	Introdução . . . . .	125
5.2	Definição dos estimadores . . . . .	127
5.3	Fisher–consistência . . . . .	128
5.4	Funções de influência parciais e variâncias assintóticas dos estimadores	133
5.5	A optimalidade da função logaritmo . . . . .	144



5.6	Estudo de simulação . . . . .	149
5.6.1	Resultados da simulação em dimensão 4 . . . . .	150
5.6.2	Resultados da simulação em dimensão 2 . . . . .	161
5.7	Exemplos com dados reais . . . . .	163
<b>6</b>	<b>Detecção de Observações Discordantes/Influentes</b>	<b>171</b>
6.1	Introdução . . . . .	171
6.2	Novas propostas . . . . .	174
6.3	Exemplos . . . . .	178
<b>7</b>	<b>Conclusões</b>	<b>187</b>
7.1	Resultados mais relevantes . . . . .	187
7.2	Trabalho futuro . . . . .	190
	<b>Apêndice A Resultados de Álgebra</b>	<b>193</b>
	<b>Apêndice B Conjuntos de dados</b>	<b>195</b>
B.1	Dados Iris (dois grupos e duas variáveis) . . . . .	195
B.2	Dados Iris Modificados (dois grupos e duas variáveis) . . . . .	197
B.3	Dados Hemofilia . . . . .	199
B.4	Dados Hemofilia Modificados . . . . .	200
B.5	Dados Variedades de Castanheiros . . . . .	201
	<b>Apêndice C Programas</b>	<b>205</b>
C.1	Programa LPP <sub>1</sub> . . . . .	205
C.2	Programa IML . . . . .	209
C.3	Programa IMB . . . . .	212
	<b>Referências Bibliográficas</b>	<b>215</b>



# Lista de Tabelas

2.1	Análise em CP para cada grupo em separado. . . . .	22
2.2	Análise em CP obtida da matriz $\mathbf{S}_P$ . . . . .	23
2.3	Análise em CPC. . . . .	24
2.4	Decomposição do teste de razão de verossimilhanças para a igualdade de $k$ matrizes de covariâncias. . . . .	31
4.1	Análise em CPC. . . . .	90
4.2	Análise robusta em CPC. . . . .	91
5.1	Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 1. . . . .	154
5.2	Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 1. . . . .	154
5.3	Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 2. . . . .	155
5.4	Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 2. . . . .	155
5.5	Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 3. . . . .	156

5.6	Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio $\lambda$ da primeira população, Modelo 3. . . . .	156
5.7	Estimação dos valores próprios de $\Sigma_1$ , Modelo 1. . . . .	157
5.8	Estimação dos valores próprios de $\Sigma_1$ , Modelo 2. . . . .	158
5.9	Estimação dos valores próprios de $\Sigma_1$ , Modelo 3. . . . .	159
5.10	Mediana e média de $\ \hat{\beta}_1 - \beta_1\ ^2$ . . . . .	162
5.11	Número de vezes, $N_\alpha$ , que o valor absoluto do ângulo entre $\beta_1$ e $\hat{\beta}_1$ é maior que $\alpha$ graus. . . . .	163
5.12	Estimação dos valores próprios de $\Sigma_1$ em dimensão 2. . . . .	163
5.13	Estimativas robustas de PI das CPC. . . . .	166
5.14	Estimativas de PP das CPC, com escala estimada através do MAD. .	166
5.15	Estimativas de PP das CPC, com escala estimada através do estimador–M. . . . .	166
5.16	Análise robusta em CP. . . . .	167
5.17	Análise clássica em CP. . . . .	167
5.18	Estimativas de PI das CPC. . . . .	168
5.19	Estimativas de PP das CPC, com escala estimada através do MAD e técnica “data direction”. . . . .	168
5.20	Estimativas de PP das CPC, com escala estimada através do estimador–M e técnica “data direction”. . . . .	168
5.21	Estimativas de PP das CPC, com escala estimada através do MAD e técnica “random direction”. . . . .	169
5.22	Estimativas de PP das CPC, com escala estimada através do estimador–M e técnica “random direction”. . . . .	169
6.1	Estimativas dos percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ . . . . .	177
6.2	Análises em CP clássicas para cada grupo em separado. . . . .	180
6.3	Análises em CP robustas para cada grupo em separado. . . . .	180
6.4	Observações detectadas como possíveis “outliers” pelo método PI e distância de Mahalanobis. . . . .	185

6.5	Observações detectadas como possíveis “outliers” pelo método PP.	. . . 185
6.6	Observações detectadas como possíveis “outliers” pelo método LPP.	. . . 185



# Lista de Figuras

1.1	Elipses correspondentes a distância de Mahalanobis constante e eixos das CP. . . . .	2
1.2	Elipses correspondentes a distância de Mahalanobis constante e eixos das CP estimados a partir das matrizes de covariâncias clássicas. . . .	3
1.3	Elipses correspondentes a distância de Mahalanobis constante e eixos das CP estimados a partir de matrizes de covariâncias robustas. . . .	4
2.1	Elipses correspondentes a distância de Mahalanobis constante e eixos das CP. . . . .	22
2.2	Elipses correspondentes a distância de Mahalanobis constante, sob CPC, e eixos das CPC. . . . .	23
4.1	Elipses correspondentes a distância de Mahalanobis constante, sob CPC, e eixos das CPC. . . . .	90
4.2	Elipses correspondentes a distância de Mahalanobis robusta constante, sob CPC, e eixos das CPC robustas. . . . .	91
4.3	Curva de enviesamento máximo de um estimador genérico, $T$ , com <i>ponto de ruptura</i> positivo e <i>função de influência</i> limitada. . . . .	119
4.4	Curva de enviesamento máximo de um estimador genérico, $T$ , com <i>ponto de ruptura</i> positivo e <i>função de influência</i> ilimitada. . . . .	120
5.1	Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 1. As linhas a preto, azul e vermelho correspondem a $C_0$ , $C_{2,0.05}$ e $C_{2,0.10}$ , respectivamente. . . . .	160

5.2	Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 2. As linhas a preto, azul e vermelho correspondem a $C_0$ , $C_{2,0.05}$ e $C_{2,0.10}$ , respectivamente.	161
5.3	Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 3. As linhas a preto, azul e vermelho correspondem a $C_0$ , $C_{2,0.05}$ e $C_{2,0.10}$ , respectivamente.	162
5.4	Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para $p = 2$ . As linhas a preto e vermelho correspondem a $C_0$ e $C_1$ , respectivamente.	164
5.5	Elipses correspondentes a distância de Mahalanobis robusta constante e eixos das CP robustas.	165
6.1	Regiões de detecção obtidas com a distância de Mahalanobis (elipse), $IML_1$ (curvas a azul) e $IMB_1$ (curvas a vermelho). Os limites de detecção a 95% são indicados com curvas contínuas e os limites a 97.5% com curvas tracejadas.	178
6.2	Elipses correspondentes a distância de Mahalanobis clássica constante e eixos das CP.	179
6.3	Elipses correspondentes a distância de Mahalanobis robusta constante e eixos das CP.	180
6.4	Detecção de observações discordantes/influente com PI clássico. As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ .	181
6.5	Detecção de observações discordantes/influente com PI robusto. As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ .	182
6.6	Detecção de observações discordantes/influente com $PP_1$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ .	182



6.7	Detecção de observações discordantes/influentes com $PP_2$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ . . . . .	183
6.8	Detecção de observações discordantes/influentes com $LPP_1$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ . . . . .	183
6.9	Detecção de observações discordantes/influentes com $LPP_2$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de $G_\lambda$ e $G_\beta$ . . . . .	184

# Capítulo 1

## Introdução

### 1.1 Objectivo desta tese

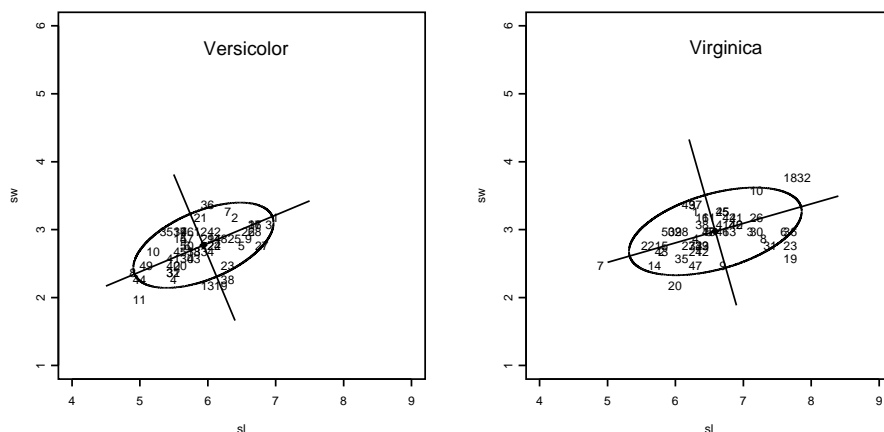
A análise em *componentes principais* (CP) é uma técnica bastante popular para analisar dados multivariados. Muitas vezes esta técnica estatística é aplicada com o intuito de “reduzir a dimensionalidade” do problema. Sem grande rigor, podemos dizer que esta análise consiste numa rotação especial do sistema de eixos original, sistema este inscrito num espaço multidimensional com dimensão igual ao número de variáveis em estudo. Rodam-se estes eixos de forma a encontrar um novo sistema de eixos, *componentes principais*, que condense tanto quanto possível a estrutura de dispersão original, mas num novo espaço com dimensão inferior ao inicial. Além disso, estas novas variáveis têm a vantagem de serem não correlacionadas.

As *componentes principais* obtêm-se com o cálculo dos vectores próprios da matriz de covariâncias, ou da matriz de correlações. Os correspondentes valores próprios medem a quantidade de informação explicada pelas *componentes principais*.

Em alguns métodos estatísticos, como é por exemplo o caso da *análise discriminante*, poderá ser importante a comparação da estrutura de covariâncias de duas ou mais populações. Uma possível via para esta comparação é a análise em *componentes principais*. Aplicando esta técnica apercebemo-nos se as matrizes de covariâncias dos vários grupos têm ou não os mesmos vectores próprios.

Para tentar ilustrar esta ideia, considerámos um conjunto de dados que é um clássico da análise multivariada. Trata-se dos dados Iris estudados por Fisher (1936) e, anteriormente, por Anderson (1935). Deste conjunto de dados seleccionámos dois grupos, *Versicolor* e *Virginica*, e as variáveis *sepal length* (*sl*) e *sepal width*

(*sw*). A bidimensionalidade das observações permite o recurso a representações gráficas. Na Figura 1.1 apresenta-se o gráfico das variáveis *sl* e *sw* para ambos os grupos. Sobrepostas às observações estão as elipses com distância de Mahalanobis (1936) amostral constante. Seguindo Flury (1988, pág. 51), considerámos  $(\mathbf{x} - \bar{\mathbf{x}}_i)^T \mathbf{S}_i^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) = 4$ , com  $\mathbf{S}_i$  a matriz de covariâncias amostral do grupo *i* e  $\bar{\mathbf{x}}_i$  a média amostral desse mesmo grupo. Incluem-se ainda os eixos principais destas elipses. Como é sabido, casos os dados sejam oriundos de um população pertencente



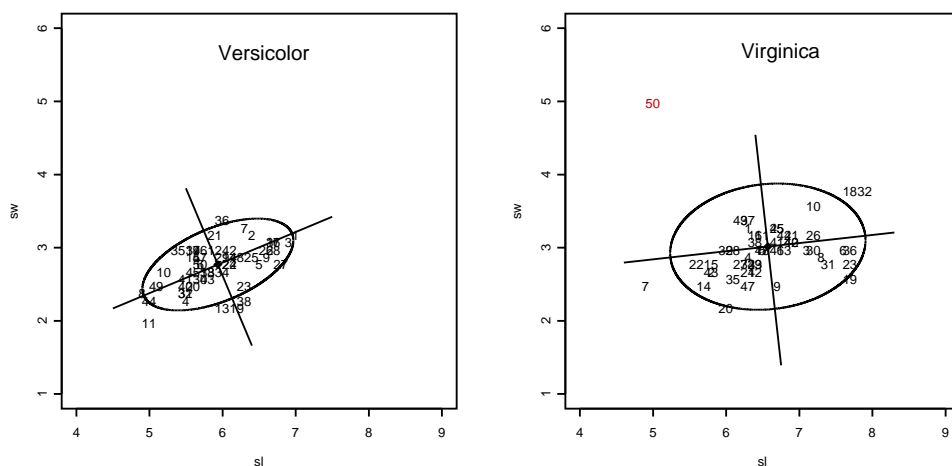
**Figura 1.1:** Elipses correspondentes a distância de Mahalanobis constante e eixos das CP.

à família de distribuições elipticamente simétricas, os eixos indicados coincidem com os eixos das *componentes principais*. Parece razoável admitir, a partir da orientação espacial dos dois sistemas de eixos principais, que as estruturas de covariâncias das variedades *Versicolor* e *Virginica* partilham os mesmos eixos principais. Assim, o que parece acontecer com a estrutura de covariâncias destes dois grupos é uma extensão das *componentes principais*. Foi isto que levou Flury (1984) a atribuir a designação de *componentes principais comuns* para modelos deste tipo, em que todas as estruturas de covariâncias partilham as mesmas *componentes principais* (vectores próprios). No entanto, é permitido que as variâncias das *componentes principais comuns* (valores próprios) sejam específicas para cada grupo.

Este modelo constitui uma boa alternativa em muitas situações reais, havendo bastantes conjuntos de dados aos quais o modelo se ajusta bem. Podemos citar, por exemplo, os dados de Jolicoeur e Mosimman (1960), Hermans e Habbema (1975), Jöreskog e Sörbom (1979), Airoldi e Hoffmann (1984), Manly (1985), Flury e Riedwyl (1988), Airoldi e Flury (1988), Krzanowski (1990), Oliveira (1995), Klingenberg *et al.* (1996), Arnold e Phillips (1999).

Do mesmo modo que no caso das *componentes principais*, podemos dizer que as *componentes principais comuns* se obtêm, de grosso modo, à custa do cálculo dos vectores próprios de matrizes de dispersão especiais. Estas matrizes de dispersão são construídas à custa de combinações lineares particulares das matrizes de covariâncias de cada grupo. Tudo isto será apresentado com maior pormenor no capítulo seguinte. O que pretendemos agora focar é o facto de todo este processo poder eventualmente ser afectado com a presença de observações anómalas ou discordantes. As observações discordantes do padrão geral dos dados afectam fortemente a estimação das matrizes de dispersão. Isto pode ter consequências graves e, analogamente ao caso das *componentes principais*, o resultado das *componentes principais comuns* pode ser influenciado. A alternativa para colmatar esta situação é o emprego de métodos robustos. Com estas técnicas estatísticas as observações discordantes não influenciam tanto a estimação das *componentes principais* ou das *componentes principais comuns*.

Uma medida muito utilizada para a detecção de observações discordantes é a distância de Mahalanobis. Se repararmos na Figura 1.1 podemos ser levados a pensar que algumas destas observações, as que estão fora dos contornos de densidade constante, poderiam ser observações discordantes. No entanto, não há nenhuma observação que apresente um comportamento muito anómalo em relação ao padrão dominante.

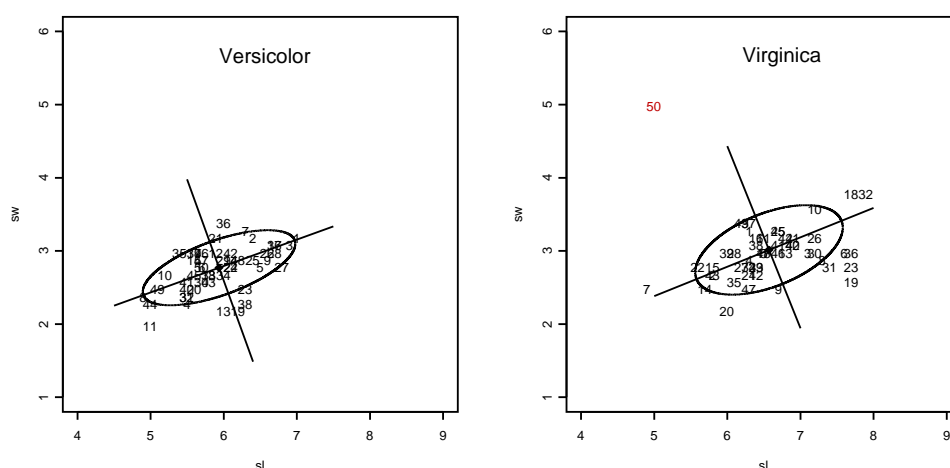


**Figura 1.2:** Elipses correspondentes a distância de Mahalanobis constante e eixos das CP estimados a partir das matrizes de covariâncias clássicas.

Será que uma observação “mais” discordante teria alguma influência na estimação das *componentes principais*?

Na tentativa de ilustrarmos a situação referida anteriormente, modificámos a última observação do grupo *Virginica*<sup>1</sup>. Como se pode observar na Figura 1.2, as *componentes principais* do segundo grupo sofreram uma rotação, no sentido horário, consequência do novo valor da observação 50. Agora, já não é tão evidente que estes grupos estão nas condições exigidas para o modelo das *componentes principais comuns*.

Por outro lado, com a estimação robusta das matrizes de covariâncias as *componentes principais* de ambos os grupos mantêm-se praticamente inalteradas, descrevendo uma situação análoga à da Figura 1.2, conforme se observa na Figura 1.3. Portanto, é bastante importante dispormos de procedimentos robustos. A comunidade



**Figura 1.3:** Elipses correspondentes a distância de Mahalanobis constante e eixos das CP estimados a partir de matrizes de covariâncias robustas.

estatística está bastante atenta a este problema. Nos últimos anos têm sido feitos esforços para desenvolver procedimentos estatísticos robustos em variadas áreas. No que respeita à análise das *componentes principais*, a utilização de métodos robustos remonta aos anos 80. Podemos destacar, dentro dos primeiros trabalhos nesta área, o trabalho de Campbell (1980), Devlin *et al.* (1981), Boente (1983), Tyler (1983), Li e Chen (1985), Critchley (1985), Boente (1987) e Pires (1990). Actualmente, este

---

<sup>1</sup>Ver Apêndice B.

tema ainda continua a ser objecto de estudo de muitos estatísticos robustos. Podemos citar, por exemplo, Croux e Haesbroeck (1999b), Croux e Haesbroeck (2000), Croux e Ruiz-Gazen (2000), Hubert *et al.* (2002) e Hubert e Rousseeuw (2002).

Por outro lado, as propostas robustas no campo das *componentes principais comuns* são muito escassas. Embora Keramidas *et al.* (1987) e, mais tarde, Schott (1991) tenham feito um breve apontamento sobre robustez para *componentes principais comuns*, o segundo autor chegou mesmo a propor um teste de hipóteses robusto, foi o trabalho de Novi Inverardi e Flury (1992) intitulado “Robust estimation of common principal components” o primeiro com esta perspectiva. Posterior a este trabalho, e que seja do nosso conhecimento, só foram publicados os trabalhos de Orellana (1999) e Boente e Orellana (2001).

Assim, o principal objectivo desta tese é a construção de novos procedimentos robustos para a análise de *componentes principais comuns*. Além disso, pretendemos também confrontar o desempenho das nossas propostas com as dos autores Novi Inverardi e Flury (1992), Orellana (1999) e Boente e Orellana (2001).

## 1.2 Metodologia utilizada

Para a construção dos novos procedimentos robustos, a metodologia utilizada foi essencialmente a abordagem infinitesimal de Hampel (1968, 1971, 1974), em especial a *função de influência*. Nenhum dos trabalhos dos outros autores seguiu esta via, apesar da *função de influência* ter um papel extremamente importante na estatística robusta.

A construção da *função de influência* para o caso de  $k$  populações foi possível após os trabalhos de Pires (1995) e Pires e Branco (2002) onde os autores introduziram as *funções de influência parciais*, definindo uma extensão do conceito de *função de influência* para problemas com  $k$  grupos.

Para apresentarmos o trabalho desenvolvido, estruturamos esta tese da forma que se segue.

Nos Capítulos 2 e 3 apresentamos resultados de vários autores que consideramos importantes para a exposição dos capítulos seguintes.

No Capítulo 2 apresentamos alguns modelos para estruturas de covariâncias. Damos especial destaque ao modelo das *componentes principais comuns*, indicando as equações que permitem a estimação de máxima verosimilhança dos parâmetros deste

modelo. Além disso, referem-se algumas propriedades destes estimadores, assim como a estatística do teste de razão de verosimilhanças. É feita também uma ilustração da aplicabilidade deste modelo.

No Capítulo 3 apresentamos algumas noções de estatística robusta, caso univariado e multivariado, dando especial destaque à abordagem de Hampel. Faz-se ainda a descrição dos estimadores robustos que são utilizados/referidos nos capítulos seguintes.

No Capítulo 4 constroem-se as *funções de influência parciais* para as *componentes principais comuns*. Estas funções são uma generalização da *função de influência* das *componentes principais* indicada, por exemplo, nos trabalhos de Critchley (1985), Pires (1990) e Croux e Haesbroeck (2000). As *funções de influência parciais* foram derivadas a partir das equações de máxima verosimilhança de Flury (1984), com a substituição dos parâmetros desconhecidos por funcionais robustos.

Para além do método da máxima verosimilhança, também é possível estimar as *componentes principais* com técnicas de “projection–pursuit” (Li e Chen, 1985). Boente e Orellana (2001), aplicaram esta técnica para estimar as *componentes principais comuns*. Para este método de estimação, derivámos as *funções de influência parciais* dos funcionais relacionados com os estimadores de “projection–pursuit”. Os nossos resultados, generalizam o trabalho de Croux e Ruiz–Gazen (2000) para  $k$  populações. Além disto, para ambos os casos, máxima verosimilhança e “projection–pursuit”, deduzimos as variâncias assintóticas dos estimadores equivalentes.

No Capítulo 5 são propostos novos estimadores “projection–pursuit”. Estes são construídos aplicando uma função geral aos estimadores “projection–pursuit” introduzidos no Capítulo 4, e designados por estimadores “projection–pursuit” gerais. São derivadas as *funções de influência parciais* destes novos funcionais e deduzidas as variâncias assintóticas dos estimadores equivalentes. Quando a função geral é a função logarítmica o estimador “projection–pursuit” resultante é o mais eficiente para a estimação dos vectores próprios comuns. Foi possível mostrar este resultado para o caso particular de matrizes de covariâncias que verifiquem o modelo *proporcional*. Este modelo é mais restrito que o modelo das *componentes principais comuns* porque exige que a razão entre as variâncias dos vectores próprios comuns de um grupo para outro seja sempre a mesma constante. Para além disto, foi feito um estudo de Monte Carlo e duas aplicações a dados reais para poder comparar os resultados da estimação das *componentes principais comuns* com os estimadores aqui sugeridos e os de Novi Inverardi e Flury (1992), Orellana (1999) e Boente e Orellana (2001).

No Capítulo 6 são propostas duas novas medidas para a detecção de observações

discordantes/influentes na estimação das *componentes principais comuns*. Estas medidas foram construídas a partir das *funções de influência parciais* dos funcionais valores próprios e vectores próprios comuns. As suas eficácias foram avaliadas com a aplicação a dois conjuntos de dados reais, os dados de Hermans e Habbema (1975) e os de Oliveira (1995). Foram confrontadas as nossas propostas com a de Rousseeuw e van Zomeren (1990) e a distância de Mahalanobis clássica.

Por último, no Capítulo 7 apresentamos as conclusões do nosso trabalho e alguns tópicos para eventuais trabalhos futuros.





## Capítulo 2

# Estruturas de Covariâncias

### 2.1 Introdução

Está no âmbito deste capítulo apresentar alguns modelos relativos a estruturas de covariâncias. Em primeiro lugar, e a título de motivação para o tema desta tese, apresentamos um modelo bastante conhecido e utilizado na prática, trata-se do modelo das *componentes principais*. No entanto, este modelo pode ser utilizado em problemas com apenas um grupo (população), e é fácil em análise multivariada defrontarmo-nos com situações onde é necessário analisar mais de um grupo. Assim, na Secção 2 iremos apresentar algumas estruturas de covariâncias para o caso de  $k$  grupos dando-se realce a certas questões que surgem no processo da estimação paramétrica. Em particular, dá-se destaque ao modelo que é objecto desta tese, o modelo das *componentes principais comuns* (CPC), proposto por Flury (1984). Além da definição e ilustração do modelo das CPC, apresentam-se os resultados obtidos por Flury que serão posteriormente utilizados/referidos nesta tese.

Na Subsecção 3.1 indicamos os resultados de Flury (1984, 1986b) que possibilitam efectuar procedimentos inferenciais nos estimadores dos eixos principais comuns (vectores próprios comuns) e das variâncias dos eixos para cada grupo (valores próprios), isto sob a validade do modelo das CPC em populações normais multivariadas. Algumas das propriedades dos estimadores dos parâmetros deste modelo encontram-se na Subsecção 3.2. A adequabilidade do modelo CPC foi ilustrada na Subsecção 3.3 com um exemplo real.

Por fim, reservou-se a última subsecção deste capítulo, a Subsecção 3.4, para apresentar uma resenha bibliográfica do tópico *componentes principais comuns*.

## 2.2 Algumas estruturas de covariâncias

Uma estrutura de covariâncias com grande aplicabilidade na prática é a subjacente à análise em *componentes principais* (Pearson, 1901). As *componentes principais* são uma das técnicas da estatística multivariada clássica mais conhecida e com aplicação em diferentes áreas científicas. Frequentemente deparamos com a necessidade de:

1. transformar um vector aleatório de  $p$  variáveis  $\mathbf{X} = (X_1, \dots, X_p)^T$  correlacionadas noutro com o mesmo número de variáveis não correlacionadas, que designaremos por  $\mathbf{U}$ ;
2. conhecer as combinações lineares de  $\mathbf{X}$  que apresentam grande ou pequena variabilidade;
3. reduzir a dimensionalidade do problema para  $q < p$ .

Estes objectivos podem ser alcançados com o uso da técnica estatística das *componentes principais*.

A sua grande popularidade deve-se em parte à atractiva interpretação geométrica destas novas variáveis aleatórias. Os elementos de  $\mathbf{U}$  não são mais do que um novo sistema de eixos construído com uma determinada rotação do sistema original com eixos  $(X_1, \dots, X_p)$ . Algebricamente, as *componentes principais* são as combinações lineares ortonormalizadas, para eliminar problemas de indeterminação, dos elementos de  $\mathbf{X}$  com variância máxima. Temos assim a seguinte definição:

### Definição 2.1 (componentes principais populacionais, Pearson, 1901)

Seja  $\mathbf{X} = (X_1, \dots, X_p)^T$  um vector aleatório para o qual se assume que existem todos os momentos de segunda ordem, e seja  $\Sigma \in PDS(p)$ <sup>1</sup> a matriz de covariâncias da população  $\mathbf{X}$ . As variáveis aleatórias  $(U_1, \dots, U_p)$  obtidas através de combinações lineares dos elementos de  $\mathbf{X}$ , i.e.,  $\mathbf{U} = \beta^T \mathbf{X}$  com  $\beta_k^T \beta_k = 1$ , que são não correlacionadas entre si e possuem sucessivamente variância máxima são conhecidas por *componentes principais* de  $\mathbf{X}$ . Isto é, a primeira *componente principal*,  $U_1 = \beta_1^T \mathbf{X}$ , é a função linear de  $\mathbf{X}$  que tem a maior variância. Por sua vez, a segunda *componente principal* será uma função linear de  $\mathbf{X}$  não correlacionada com a primeira,  $U_2 = \beta_2^T \mathbf{X}$ , que atinge a segunda maior variância. Este procedimento é sucessivamente implementado até se encontrarem as  $p$  variáveis aleatórias não correlacionadas  $\mathbf{U}$ .

---

<sup>1</sup>Denota-se por  $PDS(p)$  o espaço das matrizes reais, simétricas e definidas positivas de ordem  $p$ .

Como é sabido, quando se escolhem para colunas de  $\beta$  (matriz rotação) os vectores próprios normalizados de  $\Sigma$ , correspondentes aos valores próprios ordenados decrescentemente, obtêm-se os eixos das *componentes principais*. Mesmo no caso da multiplicidade dos valores próprios ser superior a um, digamos por exemplo  $r$ , para a construção da matriz  $\beta$  é apenas necessário estabelecer alguma convenção para a selecção de  $r$  vectores ortonormados desse subespaço.

A matriz de covariâncias de  $\mathbf{U}$  é  $\text{COV}(\mathbf{U}) = \beta^T \Sigma \beta = \Lambda$ , com  $\Lambda$  diagonal (os valores da diagonal são os valores próprios). A diagonalidade de  $\Lambda$  confirma a ausência de correlação entre os pares destas novas variáveis aleatórias. Assim, quando a decomposição espectral da matriz de covariâncias de uma população  $\mathbf{X}$ ,  $\Sigma$ , é feita com a matriz ortogonal construída com os seus vectores próprios, na matriz diagonal resultante temos as variâncias das *componentes principais*.

No entanto, esta técnica estatística desenvolve-se no contexto de uma população. Como é sabido, são muitos os problemas da análise multivariada onde temos acesso a observações das variáveis em estudo em  $k$  grupos ou populações. Este é o caso da *análise discriminante* (Fisher, 1936) ou da análise de *variáveis canónicas* (Hotelling, 1936). Nestas e noutras situações o conhecimento da estrutura de covariâncias é preponderante para prosseguir a análise estatística.

Nas análises estatísticas em problemas com  $k$  populações, é muito comum a imposição da mesma estrutura de covariâncias para todas as populações. No entanto, este pressuposto é muitas vezes violado na prática, o que se pode confirmar através de procedimentos estatísticos. Por outro lado, a análise estatística com a hipótese de matrizes de covariâncias arbitrárias também pode criar problemas devido ao número excessivo de parâmetros. O acréscimo de  $\left(\frac{(k-1)p(p+1)}{2}\right)$  parâmetros, em relação à situação anterior, faz com que esta situação seja frequentemente evitada. Até porque, e sempre que possível, devemos respeitar o princípio da parcimónia.

Na realidade, normalmente é muito mais realista admitirmos uma estrutura de covariâncias construída à custa de um compromisso entre as situações referidas, isto é, considerarmos que as covariâncias não são iguais em todos os grupos mas também não são arbitrariamente diferentes. O que poderá eventualmente existir é alguma estrutura de covariâncias comum a todos os grupos, como por exemplo uma generalização das *componentes principais*. Vários autores debruçaram-se sobre esta questão, dada a sua importância para a análise estatística, desenvolvendo técnicas gráficas e/ou analíticas. Em 1960, Jolicoeur e Mosimann, e três anos depois de novo Jolicoeur, trabalhando com *componentes principais* em dois grupos, aperceberam-se que apesar das matrizes de covariâncias não serem iguais havia muita semelhança entre

as *componentes principais* dos grupos. No entanto, a primeira tentativa de aplicação da técnica das *componentes principais* a mais de um grupo em simultâneo foi feita por Krzanowski em 1979 no trabalho intitulado “Between-groups comparison of principal components”, voltando posteriormente o autor a este tema (Krzanowski, 1982, 1984). Para detectar diferenças entre matrizes de covariâncias ou de correlações, Campbell (1981) propôs a utilização de técnicas gráficas. Todavia, foi com os trabalhos de Flury (1984, 1986b, 1987b e 1988) que se aprofundou o estudo de vários modelos com diferentes estruturas de covariâncias comuns. Embora Manly e Rayner (1987) já tenham evidenciado a criação de uma estrutura de covariâncias encaixada com quatro níveis (igualdade, proporcionalidade, igualdade das correlações e arbitrariedade das matrizes de covariâncias), foi Flury (1988) quem desenvolveu uma análise estatística completa para o seu sistema hierárquico. A formulação dessa hierarquia é descrita em seguida.

Vamos admitir que estamos na presença de  $k$  populações. Admita-se que  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$ ,  $i = 1, \dots, k$  representa um vector aleatório com todos os momentos de ordem dois.

Denote-se por  $\Sigma_i \in PSD(p)$  a matriz de covariâncias da população  $i$ . O sistema hierárquico proposto por Flury, onde o grau de semelhança vai sucessivamente enfraquecendo, é composto pelos cinco níveis (modelos) que se seguem:

**Nível 1:** Igualdade das  $k$  matrizes de covariâncias,

$$\Sigma_1 = \Sigma_2 = \dots = \Sigma_k . \quad (2.1)$$

**Nível 2:** Proporcionalidade entre as matrizes  $\Sigma_i$ , isto é,

$$\Sigma_i = \rho_i \Sigma_1 , \text{ para } i = 1, \dots, k \text{ e } \rho_1 = 1 . \quad (2.2)$$

**Nível 3:** Componentes principais comuns (CPC),

$$\Sigma_i = \beta \Lambda_i \beta^T , \quad i = 1, \dots, k , \quad (2.3)$$

onde  $\beta$  é a matriz ortogonal  $(p \times p)$  formada com os vectores próprios comuns a todas as matrizes  $\Sigma_i$ , e  $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$  são matrizes diagonais construídas com os valores próprios de cada grupo.

**Nível 4:** Componentes principais comuns parciais (CPCP( $q$ )). Seja  $q < p$  e,

$$\beta^{(i)} = (\beta_1, \dots, \beta_q, \beta_{q+1}^{(i)}, \dots, \beta_p^{(i)}) , \quad (2.4)$$

onde  $(\beta_1, \dots, \beta_q)$  são os vectores próprios comuns a todas as matrizes  $\Sigma_i$  enquanto que os restantes  $(\beta_{q+1}^{(i)}, \dots, \beta_p^{(i)})$  podem ser específicos de cada matriz. Temos:

$$\Sigma_i = \beta^{(i)} \Lambda_i \beta^{(i)T}. \quad (2.5)$$

**Nível 5:** As  $k$  matrizes de covariâncias são arbitrárias.

Como o objectivo desta tese é explorar a estrutura do Nível 3, nesta fase tecem-se algumas considerações acerca dos restantes níveis, sendo as CPC alvo de um estudo mais detalhado.

O problema da estimação, sem considerar a média, dos  $(\frac{p(p-1)}{2} + p)$  parâmetros do Nível 1, é equivalente a considerar um problema de *componentes principais* com uma matriz de covariâncias comum, por exemplo a matriz de covariâncias combinada.

A estrutura estabelecida no nível seguinte, mais forte que a das *componentes principais comuns*, necessita da estimação de mais  $(k-1)$  parâmetros que o nível anterior. Mas, como estipula uma situação mais realista e com maior ocorrência na prática que o Nível 1, esta situação foi e continua a ser objecto de estudo de vários autores. Esta constatação pode ser confirmada pelos trabalhos dos autores Federer (1951), Anderson (1958, pág. 270), Khatri (1967), Pillai *et al.* (1969), Kim (1971), Rao (1983), Guttman *et al.* (1983), Flury (1986a), Manly e Rainer (1987), Eriksen (1987) e Flury (1988, Capítulo 5), entre outros.

Quanto ao Nível 4, este surge quando o interesse está centrado no grupo das componentes que concentram maior variabilidade (variância), digamos as  $q$  componentes comuns aos  $k$  grupos podendo as restantes  $(p-q)$  diferir com o grupo. Havendo  $k$  grupos, e tal como já se referiu para o caso de um grupo, pode ser importante a redução paramétrica. Para se alcançar este objectivo procede-se à ordenação dos vectores próprios, sendo os comuns etiquetados com a sequência de inteiros de 1 a  $q$ . Quando  $q = p-1$  ou  $q = p$  estamos no caso das *componentes principais comuns*. Quando  $q = 0$ , caímos no nível seguinte. Note-se ainda que para os restantes valores de  $q$  não existe apenas um modelo de CPCP( $q$ ), mas sim uma família com modelos encaixados, pois o modelo com  $q$  componentes comuns implica o modelo com menos uma componente comum. Para este nível é necessário estimar  $kp$  parâmetros referentes às  $k$  matrizes diagonais  $\Lambda_i$ ,  $p(p-1)/2 - (p-q)(p-q-1)/2$  parâmetros relativos aos  $q$  vectores próprios comuns, e ainda os  $k(p-q)(p-q-1)/2$  correspondentes aos  $(p-q)$  vectores próprios específicos de cada grupo, perfazendo no total  $(kp + \frac{p(p-1)}{2} + \frac{(k-1)(p-q)(p-q-1)}{2})$  parâmetros.

Por último, no Nível 5 o número de parâmetro eleva-se para  $(\frac{kp(p-1)}{2} + kp)$ , mas esta questão é muitas vezes contornada optando-se por estudar os grupos em separado.

## 2.3 As componentes principais comuns

Nesta secção pretende-se apresentar, com maior destaque, o modelo do Nível 3. Este modelo, enquanto que por um lado exige a existência de um sistema de eixos comum para as  $k$  populações (vectores próprios comuns =  $\beta$ ), por outro permite que as variâncias associadas (valores próprios =  $\Lambda_i$ ) difiram com a população. No fundo, o problema apresentado retrata uma técnica semelhante às *componentes principais*, mas para o caso de várias populações. Daí que Flury (1984) o tenha rotulado por modelo das *componentes principais comuns*, designação por que é conhecido na literatura estatística.

Um problema que se pode colocar é o da aplicabilidade desta técnica estatística. Será que é sempre possível considerar um novo sistema de eixos comum a todas as populações e que goze de características semelhantes ao do sistema de eixos das *componentes principais*? Qual ou quais as condições que permitem a existência de uma matriz ortogonal, por exemplo  $\beta$ , que faz com que o resultado de  $\beta^T \Sigma_i \beta$  seja sempre uma matriz  $\Lambda_i$  diagonal, para todas as populações?

Veja-se o resultado seguinte.

### Teorema 2.1 (diagonalização simultânea de $k$ matrizes)

- (i) Se as matrizes  $(\mathbf{A}_1, \dots, \mathbf{A}_k)$  de ordem  $p$  são simultaneamente diagonalizáveis, então elas são comutáveis aos pares, isto é,  $\mathbf{A}_s \mathbf{A}_i = \mathbf{A}_i \mathbf{A}_s$  para  $s > i$ .
- (ii) Se as  $k$  matrizes são simétricas esta condição é necessária e suficiente, e a diagonalização simultânea é ortogonal.

**Dem:** A técnica seguida nesta demonstração encontra-se em Harville (1997, Capítulo 21).

(i) **Condição necessária:** Vamos supor que existe uma matriz não singular de ordem  $p$ ,  $\mathbf{Q}$ , tal que  $\mathbf{Q}^{-1} \mathbf{A}_1 \mathbf{Q} = \mathbf{D}_1, \dots, \mathbf{Q}^{-1} \mathbf{A}_k \mathbf{Q} = \mathbf{D}_k$  onde  $(\mathbf{D}_1, \dots, \mathbf{D}_k)$  são matrizes diagonais. Assim, para  $s \neq i$ , temos que

$$\begin{aligned} \mathbf{Q}^{-1} \mathbf{A}_s \mathbf{A}_i \mathbf{Q} &= \mathbf{Q}^{-1} \mathbf{A}_s \mathbf{Q} \mathbf{Q}^{-1} \mathbf{A}_i \mathbf{Q} = \\ &= \mathbf{D}_s \mathbf{D}_i = \mathbf{D}_i \mathbf{D}_s = \\ &= \mathbf{Q}^{-1} \mathbf{A}_i \mathbf{Q} \mathbf{Q}^{-1} \mathbf{A}_s \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A}_i \mathbf{A}_s \mathbf{Q}, \end{aligned}$$

o que implica

$$\mathbf{A}_i \mathbf{A}_s = \mathbf{Q} (\mathbf{Q}^{-1} \mathbf{A}_s \mathbf{A}_i \mathbf{Q}) \mathbf{Q}^{-1} = \mathbf{Q} (\mathbf{Q}^{-1} \mathbf{A}_i \mathbf{A}_s \mathbf{Q}) \mathbf{Q}^{-1} = \mathbf{A}_s \mathbf{A}_i.$$

**(ii) Condição necessária e suficiente:** Basta demonstrar que a condição é suficiente pois a demonstração da condição necessária é análoga a (i). A prova processa-se pelo método de indução matemática.

Para  $k = 1$ , esta situação está verificada, pois toda a matriz simétrica admite uma decomposição espectral com uma matriz ortogonal.

Para  $k > 1$ , suponhamos que as matrizes simétricas  $\mathbf{A}_1, \dots, \mathbf{A}_{k-1}$  são comutáveis aos pares e podem ser diagonalizáveis com uma matriz ortogonal. Sejam  $(\lambda_1, \dots, \lambda_r)$  os valores próprios distintos de  $\mathbf{A}_k$  onde  $(\nu_1, \dots, \nu_r)$  representam as suas respectivas multiplicidades.

Considere-se  $\mathbf{Q}_j$  a matriz  $(p \times \nu_j)$  formada com os vectores próprios de  $\mathbf{A}_k$ , correspondentes ao valor próprio  $\lambda_j$ , e cujas colunas formam uma base ortonormada do espaço linear nulo  $(\nu_j$  dimensional)  $\mathcal{N}(\mathbf{A}_k - \lambda_j \mathbf{I})$  para  $j = 1, \dots, k$ . Tome-se  $\mathbf{Q} = (\mathbf{Q}_1, \dots, \mathbf{Q}_r)$ . Como  $\mathbf{A}_k$  é diagonalizável, porque é simétrica, tem-se que  $\sum_{j=1}^r \nu_j = p$  pelo Resultado 1 do Apêndice A. Assim,  $\mathbf{Q}$  é uma matriz ortogonal de ordem  $p$ .

A decomposição espectral de  $\mathbf{A}_k$  com essa matriz é

$$\mathbf{Q}^T \mathbf{A}_k \mathbf{Q} = \text{diag}(\lambda_1 \mathbf{I}_{\nu_1}, \dots, \lambda_r \mathbf{I}_{\nu_r}).$$

Como  $\mathbf{A}_k$  comuta com as restantes  $(k - 1)$  matrizes, para  $i \neq j$ , tem-se que

$$\mathbf{A}_k (\mathbf{A}_i \mathbf{Q}_j) = \mathbf{A}_i (\mathbf{A}_k \mathbf{Q}_j) = \lambda_j \mathbf{A}_i \mathbf{Q}_j,$$

resultando a igualdade  $(\mathbf{A}_k - \lambda_j \mathbf{I}) \mathbf{A}_i \mathbf{Q}_j = \mathbf{0}_p$ .

O Resultado 2 do Apêndice A conjuntamente com o facto das colunas da matriz  $\mathbf{Q}_j$  constituírem uma base do espaço nulo  $\mathcal{N}(\mathbf{A}_k - \lambda_j \mathbf{I})$ , origina a existência de uma matriz de ordem  $\nu_j$ , por exemplo  $\mathbf{B}_{ij}$ , que pré-multiplicada por  $\mathbf{Q}_j$  iguala  $\mathbf{A}_i \mathbf{Q}_j$ ,  $i = 1, \dots, k - 1$ . Como  $\mathbf{B}_{ij} = \mathbf{I}_{\nu_j} \mathbf{B}_{ij}$ , e por  $\mathbf{I}_{\nu_j}$  poder ser escrita à custa de  $\mathbf{Q}_j$ , resulta  $\mathbf{B}_{ij} = \mathbf{Q}_j^T \mathbf{A}_i \mathbf{Q}_j$  o que prova a simetria.

Pode também verificar-se que as matrizes deste tipo (de ordem  $\nu_j$ ) são comutáveis aos pares, pois para  $s > i$  e considerando por exemplo  $\mathbf{B}_{ij}$  e  $\mathbf{B}_{sj}$ , tem-se

$$\mathbf{B}_{sj} \mathbf{B}_{ij} = \mathbf{Q}_j^T \mathbf{A}_s \mathbf{Q}_j \mathbf{B}_{ij} = \mathbf{Q}_j^T \mathbf{A}_s \mathbf{A}_i \mathbf{Q}_j = \mathbf{Q}_j^T \mathbf{A}_i \mathbf{A}_s \mathbf{Q}_j = \mathbf{Q}_j^T \mathbf{A}_i \mathbf{Q}_j \mathbf{B}_{sj} = \mathbf{B}_{ij} \mathbf{B}_{sj}.$$



Assim, pela hipótese de indução as  $(k - 1)$  matrizes  $(\mathbf{B}_{1j}, \dots, \mathbf{B}_{k-1,j})$  podem ser diagonalizáveis com uma matriz ortogonal. Designando por  $\mathbf{W}_j$  essa matriz ortogonal vem  $\mathbf{W}_j^T \mathbf{B}_{ij} \mathbf{W}_j = \mathbf{D}_{ij}$ , onde  $\mathbf{D}_{ij}$  é uma matriz diagonal. Seja  $\mathbf{W}$  a matriz ortogonal definida como  $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_r)$  e  $\mathbf{P}$  a matriz obtida pela sua pré-multiplicação com  $\mathbf{Q}$  que também é ortogonal.

A decomposição espectral de  $\mathbf{A}_i$  com  $\mathbf{Q}$  é

$$\begin{aligned} \mathbf{Q}^T \mathbf{A}_i \mathbf{Q} &= \mathbf{Q}^T (\mathbf{A}_i \mathbf{Q}_1, \dots, \mathbf{A}_i \mathbf{Q}_r) = \\ &= \mathbf{Q}^T (\mathbf{Q}_1 \mathbf{B}_{i1}, \dots, \mathbf{Q}_r \mathbf{B}_{ir}). \end{aligned} \quad (2.6)$$

O facto de  $\mathbf{Q}$  ser ortogonal faz com que o último membro desta equação seja uma matriz diagonal construída com os elementos  $(\mathbf{B}_{i1}, \dots, \mathbf{B}_{ik})$ .

Assim, temos que

$$\begin{aligned} \mathbf{P}^T \mathbf{A}_i \mathbf{P} &= \mathbf{W}^T \mathbf{Q}^T \mathbf{A}_i \mathbf{Q} \mathbf{W} = \text{diag}(\mathbf{W}_1^T \mathbf{B}_{i1} \mathbf{W}_1, \dots, \mathbf{W}_r^T \mathbf{B}_{ir} \mathbf{W}_r) = \\ &= \text{diag}(\mathbf{D}_{i1}, \dots, \mathbf{D}_{ir}). \end{aligned}$$

A decomposição de  $\mathbf{A}_k$  é

$$\begin{aligned} \mathbf{P}^T \mathbf{A}_k \mathbf{P} &= \mathbf{W}^T \mathbf{Q}^T \mathbf{A}_k \mathbf{Q} \mathbf{W} = \text{diag}(\lambda_1 \mathbf{W}_1^T \mathbf{W}_1, \dots, \lambda_r \mathbf{W}_r^T \mathbf{W}_r) = \\ &= \text{diag}(\lambda_1 \mathbf{I}_{\nu_1}, \dots, \lambda_r \mathbf{I}_{\nu_r}), \end{aligned} \quad (2.7)$$

o que permite concluir que as matrizes  $(\mathbf{A}_1, \dots, \mathbf{A}_k)$  são mutuamente diagonalizáveis com a mesma matriz ortogonal  $\mathbf{Q}$ , o que conclui a prova.  $\square$

Assim, o modelo das CPC é uma boa opção quando estamos perante populações com matrizes de covariâncias comutáveis aos pares, pois a exigência da diagonalização simultânea a tal conduz. Nesta situação as novas variáveis aleatórias (simultaneamente transformadas)  $\mathbf{U}_i = \boldsymbol{\beta}^T \mathbf{X}_i$ , referenciadas no novo sistema de eixos (comum às  $k$  populações) são designadas por *componentes principais comuns* devido à analogia com as *componentes principais*. Convém notar que não existe nenhuma ordenação particular para as colunas da matriz  $\boldsymbol{\beta}$ , contrariamente ao que acontece no caso de  $k = 1$ , uma vez que a ordem dos elementos nas  $k$  matrizes diagonais não é necessariamente a mesma. Normalmente, para se ordenar as colunas desta matriz a convenção que se utiliza é  $(\boldsymbol{\beta}_1^T \boldsymbol{\Sigma}_1 \boldsymbol{\beta}_1 > \boldsymbol{\beta}_2^T \boldsymbol{\Sigma}_1 \boldsymbol{\beta}_2 > \dots > \boldsymbol{\beta}_p^T \boldsymbol{\Sigma}_1 \boldsymbol{\beta}_p)$ , isto no caso de  $\boldsymbol{\Sigma}_1$  ter  $p$  valores próprios distintos.

O facto da rotação ser ortogonal e a mesma para as  $k$  populações proporciona um ganho na estimação paramétrica, no que refere à matriz  $\boldsymbol{\beta}$ , sendo apenas necessário estimar  $(\frac{p(p-1)}{2})$  parâmetros. Comparando com o Nível 5 temos um número de parâmetros  $k$  vezes menor. No total, considerando também os valores próprios, existem  $(\frac{p(p-1)}{2} + kp)$  parâmetros a estimar.

### 2.3.1 Estimação dos parâmetros por máxima verosimilhança

Nesta subsecção apresentaremos resumidamente as equações de verosimilhança construídas por Flury (1984) e que permitem estimar, com o recurso a um processo iterativo, os parâmetros do modelo das CPC.

Para isso, vamos supor que  $\mathbf{X}_i$  ( $i = 1, \dots, k$ ) são vectores aleatórios de  $\mathbb{R}^p$  distribuídos segundo uma normal- $p$  variada, com matrizes de covariâncias  $\Sigma_i \in PDS(p)$ . Além disto, supõe-se que as  $k$  matrizes  $\Sigma_i$  verificam o modelo das CPC, apresentado em (2.3).

Para amostras de dimensão  $n_i$ , considere-se  $\mathbf{S}_i$  o estimador centrado usual da matriz de covariâncias  $\Sigma_i$ , isto é,  $\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i) (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^T$  com  $i = 1, \dots, k$ . Assume-se ainda que em todos os grupos a dimensão das amostras é superior ao número de variáveis, i.e.,  $\min_{i=1, \dots, k} n_i \geq p + 1$ . Com isto resulta que para as  $k$  matrizes independentes se verifica  $(n_i - 1)\mathbf{S}_i \sim W_p((n_i - 1), \Sigma_i)$ , como se pode ver, por exemplo, em Muirhead (1982, pág. 85).

Considere-se  $N = \sum_{i=1}^k n_i$  o número total de observações, e  $\tau_i = \frac{n_i}{N}$  a proporção de elementos do  $i$ -ésimo grupo.

A função de verosimilhança conjunta de  $(\Sigma_1, \dots, \Sigma_k)$ , condicional a  $(\mathbf{S}_1, \dots, \mathbf{S}_k)$ , vem

$$L(\Sigma_1, \dots, \Sigma_k | \mathbf{S}_1, \dots, \mathbf{S}_k) = c \times \prod_{i=1}^k \text{etr} \left( -\frac{(n_i - 1)}{2} \Sigma_i^{-1} \mathbf{S}_i \right) |\Sigma_i|^{-(n_i - 1)/2}, \quad (2.8)$$

sendo  $c$  uma constante independente de  $\Sigma_i$ , e  $\text{etr}$  a função exponencial aplicada ao operador traço.

Maximizar a função (2.8), sob a validade de (2.3), é equivalente a minimizar a seguinte função de  $\beta$  e  $\Lambda_i$

$$g(\Sigma_1, \dots, \Sigma_k | \beta; \Lambda_1, \dots, \Lambda_k) = \sum_{i=1}^k (n_i - 1) \sum_{j=1}^p \left[ \log \lambda_{ij} + \frac{\beta_j^T \mathbf{S}_i \beta_j}{\lambda_{ij}} \right], \quad (2.9)$$

sujeita à restrição de  $\beta$  ortogonal, i.e.,  $\beta^T \beta = \mathbf{I}_p$ .

Com a aplicação dos multiplicadores de Lagrange e igualando a zero as derivadas parciais em ordem aos parâmetros a estimar, resulta que os estimadores de máxima

verossimilhança (EMV) têm que verificar

$$\boldsymbol{\beta}_j^T \mathbf{S}_i \boldsymbol{\beta}_j = \lambda_{ij} , \quad (2.10)$$

$$\boldsymbol{\beta}_m^T \left( \sum_{i=1}^k (n_i - 1) \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \mathbf{S}_i \right) \boldsymbol{\beta}_j = 0 , \quad \text{para } m \neq j , \quad (2.11)$$

$$\boldsymbol{\beta}_m^T \boldsymbol{\beta}_j = \delta_{mj} , \quad (2.12)$$

onde  $\delta_{mj} = 1$  quando  $m = j$  e é 0 no caso contrário.

Flury e Gautschi (1986) mostraram que a maximização da verossimilhança (2.8) é também equivalente à minimização da função

$$\Phi(\boldsymbol{\beta}) = \prod_{i=1}^k [\det(\text{diag}(\mathbf{F}_i)) / \det(\mathbf{F}_i)]^{(n_i-1)} , \quad (2.13)$$

com  $\boldsymbol{\beta} \in \mathcal{O}(p)$ , o grupo de matrizes ortogonais ( $p \times p$ ) e  $\mathbf{F}_i = \boldsymbol{\beta}^T \mathbf{S}_i \boldsymbol{\beta}$ . Este facto permitiu aos referidos autores a construção de um algoritmo, designado por algoritmo F-G, com o qual se pode determinar numericamente uma possível solução de máxima verossimilhança e que será referido com maior pormenor na Subsecção 3.4. Para o caso de 2 grupos em dimensão 2, os autores indicam condições aproximadas para a existência de um único ponto de mínimo. Este resultado está relacionado com o valor da excentricidade das matrizes, i.e., o quociente entre o maior e o menor valor próprio. Para problemas com mais de duas populações, os autores também indicam quais as situações onde pode ocorrer mais de um mínimo local. Isto apresenta-se como consequência da grande excentricidade das matrizes, ou quando o ângulo entre dois vectores próprios de grupos distintos é próximo de  $\frac{\pi}{4}$  (situação afastada da validade do modelo CPC). No entanto, para os casos onde existe uma prévia suspeita a favor de  $k$  mínimos locais, Flury e Gautschi (1986) recomendam o uso do algoritmo F-G  $k$  vezes. Nessa situação, a solução inicial para a  $i$ -ésima utilização do algoritmo F-G deve ser a matriz construída com os vectores próprios da  $\mathbf{S}_i$ . Quando as  $k$  soluções são semelhantes parece razoável admitir-se a existência de um único mínimo global. Em Flury (1988, págs. 194–200) encontram-se mais pormenores sobre esta questão.

As equações de verossimilhança (2.10), (2.11) e (2.12) têm uma interpretação atractiva. Se designarmos por  $\mathbf{A}_{mj}^{(i)}$  a matriz simétrica apresentada em (2.11) dentro de parêntesis, de  $\boldsymbol{\beta}_m^T \mathbf{A}_{mj}^{(i)} \boldsymbol{\beta}_j = 0$  é claro que  $\boldsymbol{\beta}_m$  e  $\boldsymbol{\beta}_j$  são vectores característicos de  $\mathbf{A}_{mj}^{(i)}$ . Por isso, o sistema de equações (2.11) pode ser encarado como um sistema generalizado de equações características. Caso todas as matrizes  $\mathbf{A}_{mj}^{(i)}$  sejam idênticas, isto para  $m < j$ , este sistema conduz aos valores próprios da matriz comum. No

entanto, isto não acontece, pois para cada par  $(m, j)$  a matriz  $\mathbf{A}_{mj}^{(i)}$  é construída com uma combinação linear diferente de  $(n_i - 1)\mathbf{S}_i$ . Os valores dos pesos das matrizes  $(n_i - 1)\mathbf{S}_i$  são tanto menores quanto mais próximas sejam as variâncias das duas combinações lineares  $\mathbf{U}_{im} = \boldsymbol{\beta}_m^T \mathbf{X}_i$  e  $\mathbf{U}_{ij} = \boldsymbol{\beta}_j^T \mathbf{X}_i$ .

No caso da igualdade destas variâncias, i.e., quando  $(\lambda_{im} = \lambda_{ij})$ , a  $i$ -ésima parcela da equação  $(m, j)$  anula-se. Isto é justificado pelo facto da quantidade  $(\lambda_{im} - \lambda_{ij})$  traduzir a esfericidade de  $\mathbf{X}_i$  no subespaço gerado pelos vectores  $\boldsymbol{\beta}_m$  e  $\boldsymbol{\beta}_j$ . Neste caso, a projecção das observações no subespaço gerado pelos eixos  $\mathbf{u}_{im}$  e  $\mathbf{u}_{ij}$  é uma esfera perfeita.

Note-se ainda que o sistema de equações (2.11) não sofre qualquer alteração com a dilatação da matriz de covariâncias  $\mathbf{S}_i$ . Representando a matriz dilatada por  $\mathbf{S}_i^* = c_i \mathbf{S}_i$ , com  $c_i > 0$ , tendo em conta (2.10) é fácil verificar que a matriz  $\mathbf{A}_{mj}^{(i)}$  não é afectada com a proporcionalidade. Isto é de certa forma intuitivo, uma vez que não faria sentido que o “tamanho” da matriz  $\mathbf{S}_i$ , em termos do determinante ou do traço da matriz, influenciasse na estimação dos eixos principais comuns.

A análise em CPC pode ser interpretada como uma transformação simultânea de  $k$  vectores de  $\mathbb{R}^p$  de forma a obter novos vectores (variáveis). Estas novas variáveis também de  $\mathbb{R}^p$  devem ser o menos correlacionadas possível. Com isto queremos dizer que a matriz de covariâncias amostral das variáveis  $\mathbf{U}_i = \boldsymbol{\beta}^T \mathbf{X}_i$ , ou seja a matriz  $\mathbf{F}_i$ , deve ser o mais próximo possível da forma diagonal.

### 2.3.2 Propriedades dos estimadores

Vamos considerar que o par  $(\hat{\boldsymbol{\Lambda}}_i, \hat{\boldsymbol{\beta}})$  representa os estimadores de máxima verosimilhança de  $(\boldsymbol{\Lambda}_i, \boldsymbol{\beta})$ . A matriz de covariâncias de  $i$ -ésima população é estimada como

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\beta}}^T, \quad i = 1, \dots, k. \quad (2.14)$$

As *componentes principais comuns* amostrais definem-se por  $\mathbf{U}_i = \hat{\boldsymbol{\beta}}^T \mathbf{X}_i$ , para  $i = 1, \dots, k$ , sendo a matriz de covariâncias amostral destas novas variáveis igual a

$$\hat{\mathbf{F}}_i = \hat{\boldsymbol{\beta}}^T \mathbf{S}_i \hat{\boldsymbol{\beta}}, \quad i = 1, \dots, k. \quad (2.15)$$

Da equação (2.10) resulta também que

$$\hat{\boldsymbol{\Lambda}}_i = \text{diag}(\hat{\mathbf{F}}_i), \quad i = 1, \dots, k. \quad (2.16)$$

Como se viu na subsecção anterior, as CPC são a transformação ortogonal que diagonaliza em simultâneo  $k$  matrizes de covariâncias, quando isto é possível. Assim,

é conveniente haver um critério que avalie a apropriabilidade deste processo. Neste sentido, Flury (1984) propôs o uso de um teste de razão de verossimilhanças para avaliar a validade do modelo das CPC.

Denotando a hipótese a testar por  $H_{\text{CPC}} : \boldsymbol{\beta}^T \boldsymbol{\Sigma}_i \boldsymbol{\beta} = \boldsymbol{\Lambda}_i \quad i = 1, \dots, k$ , tendo em conta que no máximo da função de verossimilhança  $\text{tr}(\hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{S}_i) = p$  assim como os resultados (2.15) e (2.16), o quociente dos máximos das verossimilhanças, sendo o numerador restrito a  $H_{\text{CPC}}$  e o denominador irrestrito, vem

$$X^2 = -2 \log \frac{L(\hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\boldsymbol{\Sigma}}_k)}{L(\mathbf{S}_1, \dots, \mathbf{S}_k)} = \sum_{i=1}^k \log \frac{|\hat{\boldsymbol{\Sigma}}_i|}{|\mathbf{S}_i|} = \sum_{i=1}^k \log \frac{|\text{diag} \hat{\mathbf{F}}_i|}{|\hat{\mathbf{F}}_i|}. \quad (2.17)$$

Pela teoria geral dos testes de razão de verossimilhanças (Rao, 1973, págs. 418–419) segue-se que, sob a validade de  $H_{\text{CPC}}$ , a estatística de teste  $X^2$  distribui-se aproximadamente segundo um qui-quadrado com  $[(k-1)(p-1)p]/2$  graus de liberdade. Este número de graus de liberdade é facilmente obtido subtraindo o número de parâmetros a estimar no caso irrestrito (Nível 5) com o número a estimar restrito a  $H_{\text{CPC}}$ .

Flury (1986b) obteve a distribuição assymptótica dos estimadores  $\hat{\boldsymbol{\Lambda}}_i$  e  $\hat{\boldsymbol{\beta}}$ . Para apresentar a distribuição assymptótica de  $\hat{\boldsymbol{\beta}}$ , Flury (1986b, 1988, págs. 76 e 79) utilizou o operador  $\text{vec}$  (ver Apêndice A).

Os resultados principais encontram-se no teorema seguinte.

**Teorema 2.2 (distribuição assymptótica dos estimadores do modelo CPC, Flury, 1986b, 1988, págs. 74–80)** Suponha-se que são verificadas as condições do modelo (2.3) e a notação introduzida na Subsecção 3.1. Então, quando o  $\min(n_i) \rightarrow \infty$  resulta que

1. as  $pk$  variáveis  $\sqrt{(n_i - 1)} (\hat{\lambda}_{ij} - \lambda_{ij})$  têm distribuição assymptótica  $N(0, 2\lambda_{ij}^2)$ , são independentes entre si e também independentes de  $\hat{\boldsymbol{\beta}}$ ;
2. a distribuição assymptótica de  $\sqrt{(n_i - 1)} \text{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  é uma  $N(\mathbf{0}_p, \mathbf{V}_{\hat{\boldsymbol{\beta}}})$ .

A matriz de covariâncias  $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$  é uma matriz  $(p^2 \times p^2)$  e tem na entrada  $(m, m)$

a matriz  $\left( \sum_{\substack{j=1 \\ j \neq m}}^p \theta_{mj} \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T \right), (p \times p)$ , enquanto que na entrada  $(m, j)$  tem-se a

matriz  $(-\theta_{mj}\beta_j\beta_m^T)$ ,  $(p \times p)$ , com  $\theta_{mj} = \left[ \sum_{i=1}^k (\theta_{mj}^{(i)})^{-1} \right]^{-1}$  e

$$\theta_{mj}^{(i)} = \tau_i^{-1} \frac{\lambda_{im}\lambda_{ij}}{(\lambda_{im} - \lambda_{ij})^2} \quad m \neq j.$$

A demonstração destes resultados encontra-se em Flury (1986b, 1988, págs. 74–80).

### 2.3.3 Exemplo

Com esta subsecção pretendemos ilustrar a aplicação do modelo das CPC. Para isto seleccionámos um exemplo que possibilitasse recorrer a representações gráficas. Como vimos no Capítulo 1, estas representações permitem avaliar visualmente se a escolha deste modelo é ou não adequada para os dados em estudo. Este exemplo será posteriormente, nesta tese, objecto de uma análise mais detalhada.

Em Flury (1988, págs. 85–99) encontram-se outros exemplos onde o modelo CPC se mostra como uma boa alternativa para a modelação da estrutura de covariâncias dos grupos.

#### Exemplo 1 (Hemofilia)

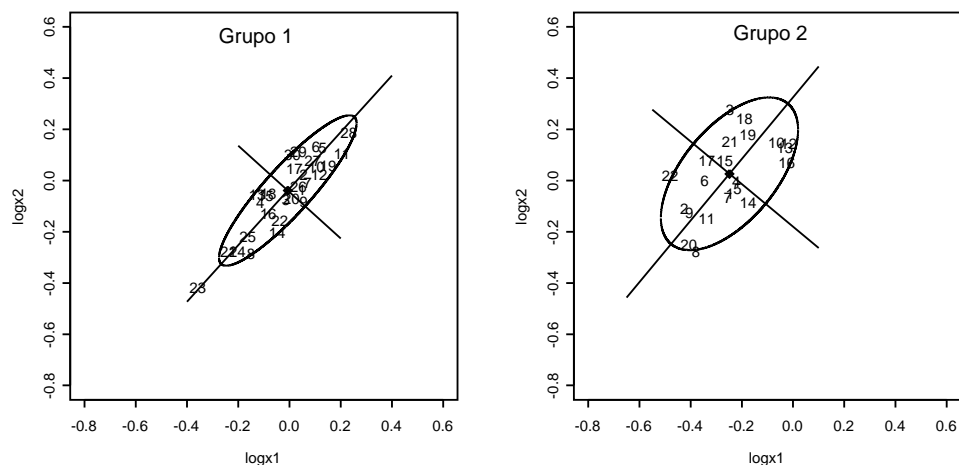
Para a ilustração do modelo das CPC seleccionámos a situação mais simples onde é possível aplicar a técnica das CPC, o caso em que se observam apenas duas variáveis em dois grupos.

Estes dados foram também estudados por outros autores mas no contexto da *análise discriminante*. Nesses estudos pretendia-se discriminar as mulheres portadoras de hemofilia das não portadoras desse gene recessivo. Podemos citar por exemplo, os autores Hermans e Habbema (1975) e Pires (1995) entre os que analisaram os dados nesse contexto. Estes estudos foram possíveis após Zimmerman *et al.* (1971) e de Bouma *et al.* (1975) concluírem que a observação de duas variáveis, com análises sanguíneas, permitia inferir sobre a capacidade de coagulação.

Neste conjunto de dados constam dois grupos, constituídos por 30 mulheres não portadoras (Grupo 1) e 22 mulheres portadoras desse gene (Grupo 2). As observações dos logaritmos (na base 10) das duas variáveis, FactorVIII-actividade ( $x_1$ ) e FactorVIII-antigene ( $x_2$ ) encontram-se no Apêndice B.

A modelação da estrutura de covariâncias é fundamental para a construção da regra discriminante. Caso a hipótese da igualdade das covariâncias não seja confirmada por estes dados, a regra discriminante quadrática, construída supondo as covariâncias diferentes, pode não estar a respeitar o princípio da parcimónia. O modelo

das CPC poderá eventualmente ser uma alternativa intermédia e servir de ponto de partida para a construção doutra regra discriminante mais simples. Na Figura



**Figura 2.1:** Elipses correspondentes a distância de Mahalanobis constante e eixos das CP.

2.1 apresentamos as observações, sobrepondo as elipses com distância de Mahalanobis amostral constante. Este tipo de representação, usado pela primeira vez por Hotelling (1933) no contexto das *componentes principais*, permite avaliar visualmente a adequação do modelo das *componentes principais comuns* aos dados. Os contornos representados correspondem a  $(\mathbf{x} - \bar{\mathbf{x}}_i)^T \mathbf{S}_i^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) = 4$ , com  $\bar{\mathbf{x}}_i$  o vector média amostral e  $\mathbf{S}_i$  a matriz de covariâncias amostral do grupo  $i = 1, 2$ . Incluem-se ainda os eixos principais dessas elipses que indicam as direcções das *componentes principais*. Da observação da Tabela 2.1 ressalta a semelhança entre as *componentes*

	<i>Grupo 1</i>		<i>Grupo 2</i>	
	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>1</sub>	CP <sub>2</sub>
valores próprios	0.036	0.002	0.030	0.008
vectors próprios				
$\log_{10} x_1$	0.671	0.741	0.639	0.769
$\log_{10} x_2$	0.741	-0.671	0.769	-0.639

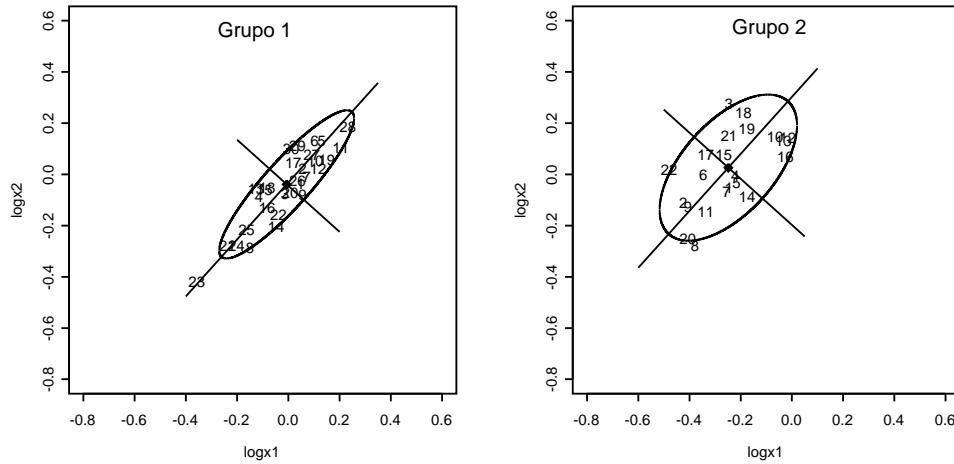
**Tabela 2.1:** Análise em CP para cada grupo em separado.

*principais* dos dois grupos. As direcções quase paralelas dos eixos das duas elipses

da Figura 2.1 também reflectem esta constatação. Além disso, se observarmos as estimativas das *componentes principais* obtidas com a matriz de covariâncias combinada (Tabela 2.2),  $\mathbf{S}_P = \frac{29S_1+21S_2}{50}$ , verificamos as semelhanças entre estes eixos e os obtidos com as análises para cada grupo separadamente, confirmando a eventual validade de  $H_{CPC}$  (Krzanowski, 1984). Assim, parece estarmos na presença

	CP <sub>1</sub>	CP <sub>2</sub>
valores próprios	0.035	0.005
<hr/>		
	vectores próprios	
$\log_{10} x_1$	0.661	0.750
$\log_{10} x_2$	0.750	-0.661

**Tabela 2.2:** Análise em CP obtida da matriz  $\mathbf{S}_P$ .



**Figura 2.2:** Elipses correspondentes a distância de Mahalanobis constante, sob CPC, e eixos das CPC.

de um conjunto de dados para o qual o modelo das CPC é realmente alternativa. Veja-se ainda na Figura 2.2 como os eixos das novas elipses calculadas através de  $(\mathbf{x} - \bar{\mathbf{x}}_i)^T \hat{\Sigma}_i (\mathbf{x} - \bar{\mathbf{x}}_i) = 4$ , com  $\hat{\Sigma}_i$  a estimativa da matriz de covariâncias sob o modelo das CPC, não estão muito rodados em relação aos da Figura 2.1. As elipses da Figura 2.2 têm, por definição, os eixos paralelos.



Para a estimação das matrizes  $\Sigma_i$ , sob a validade do modelo CPC, utilizamos o programa da autoria de Phillips (1994). Este programa encontra-se no seguinte endereço da “internet”, <http://wbar.uta.edu/software/cpc.htm>. Para além das matrizes  $\hat{\Sigma}_i$ , dos eixos principais comuns e respectivas variâncias para cada grupo, este programa também fornece o resultado do teste de razão de verosimilhanças apresentado em (2.17).

Também é possível obter as estimativas de  $(\Lambda_i, \beta)$  através do programa S-PLUS 2000. Este programa permite construir uma regra discriminante sob a validade do modelo CPC. Para isso utiliza-se o comando **discrim**, especificando que a estrutura de covariâncias satisfaz o modelo CPC, e uma das saídas deste comando é  $(\hat{\Lambda}_i, \hat{\beta})$ .

Os resultados da estimação dos vectores próprios comuns e valores próprios associados para cada grupo, são apresentados na Tabela 2.3. A adequação do modelo CPC

vectores próprios			valores próprios		
$\log_{10} x_1$	0.669	0.744	<i>Grupo 1</i>	0.036	0.002
$\log_{10} x_2$	0.744	-0.669	<i>Grupo 2</i>	0.030	0.008

**Tabela 2.3:** Análise em CPC.

para estas observações é confirmada pelo teste de ajustamento de  $H_{CPC}$ . Obteve-se para valor observado da estatística 0.072, o que confrontado com os quantis de um  $\chi^2_{(1)}$  traduz um *valor-p* da ordem dos 78%.

### 2.3.4 Revisão bibliográfica

Nesta subsecção pretendemos apresentar os trabalhos existentes sobre o tema das *componentes principais comuns*. Sobre cada trabalho será feita uma breve descrição e tecem-se também alguns comentários que pareçam relevantes.

Por ordem cronológica, temos:

**Krzanowski, W. J. (1979). Between-groups comparison of principal components.**

Este foi o primeiro trabalho onde se tentou encontrar *componentes principais* para  $k$  grupos. A ideia de Krzanowski foi criar um subconjunto  $q$ -dimensional, do espaço original  $p$ -dimensional, que fosse comum às populações. A comparação dos  $k$  sub-espacos foi feita à custa dos ângulos mínimos entre eles. Para isso, considerou para todos os  $k$  grupos só as primeiras  $q$  *componentes principais*. Representou por  $\mathbf{L}_t$

(para o grupo  $t$ ) a matriz onde a entrada  $(i, j)$  tem o valor da  $j$ -ésima componente do vector próprio correspondente ao  $i$ -ésimo valor próprio ( $t = 1, \dots, k$ ). O ângulo mínimo entre um vector arbitrário do espaço inicial,  $\mathbf{b}$ , e outro com direcção próxima do paralelo no subespaço das  $q$  componentes é dado por  $\delta = \cos^{-1}(\mathbf{b}_1^T \mathbf{L}_t^T \mathbf{L}_t \mathbf{b}_1)^{\frac{1}{2}}$ . Nesta expressão  $\mathbf{b}_1$  é o vector próprio correspondente ao maior valor próprio de  $\mathbf{H} = \sum_{t=1}^k \mathbf{L}_t^T \mathbf{L}_t$ . Isto significa que, no caso particular do vector arbitrário ser  $\mathbf{b}_1$  o ângulo alcançado é mínimo. Com  $\mathbf{b}_2$ , vector próprio associado ao segundo maior valor próprio (com direcção ortogonal à de  $\mathbf{b}_1$ ), substituído na expressão do argumento do arco coseno encontra-se o ângulo imediatamente superior ao anterior. Este processo é continuado até ao último valor (vector) próprio da matriz  $\mathbf{H}$ .

Esta abordagem apresenta problemas quando os valores próprios  $q$  e  $(q + 1)$  são próximos para um determinado grupo. Neste caso a instabilidade dos vectores próprios poderá originar ângulos entre os subespaços ( $q$ -dimensionais) demasiado grandes e conduzir erradamente à não aceitação do subespaço comum. Para o caso em que o subespaço comum parece funcionar, Krzanowski também não propôs nenhum método para a estimação das matrizes de covariâncias. No entanto, este trabalho teve o mérito de levantar o problema da comparação das *componentes principais* de vários grupos, que foi e ainda continua a ser um tema que merece a atenção da comunidade estatística.

Como aplicação o autor utilizou um conjunto de dados relacionados com os resultados obtidos por alunos venezuelanos em escolas localizadas no norte de Inglaterra. No estudo de Krzanowski entrevistaram 3 escolas e a análise foi desenvolvida a partir do resultado da observação de 8 variáveis.

**Krzanowski, W. J. (1982). Between-groups comparison of principal components-some sampling results.**

Neste trabalho Krzanowski investigou o comportamento de  $\delta$  com um processo de simulação. No delineamento da simulação incluiu diferentes valores para  $q$ ,  $p$ , considerou amostras de várias dimensões e várias estruturas de covariâncias, mas sempre para o caso de  $k = 2$ .

Os resultados da simulação permitiram a Krzanowski dar um valor limite para o ângulo  $\delta$ , ou seja, uma tolerância para o valor do ângulo mínimo.

**Flury, B. N. (1983a). Some relations between the comparison of covariance matrices and principal component analysis.**

Neste trabalho Flury tentou evidenciar a importância da comparação de várias matrizes de covariâncias. Normalmente este tema só era estudado pelos estatísticos com

a finalidade de validar, ou não, a igualdade das matrizes de covariâncias, condição necessária para alguns dos procedimentos estatísticos multivariados. Flury tentou realçar a importância da comparação de duas matrizes de covariâncias, encarando este problema como uma generalização da análise em *componentes principais* para mais grupos.

Baseado no teste de Roy (1957, pág. 28), teste de união-intersecção, para a igualdade de matrizes de dispersão, Flury sugeriu a utilização da matriz  $\mathbf{A} = \mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_2$  para uma comparação mais geral. A análise das combinações lineares dos vectores próprios desta matriz é útil, segundo este autor, para casos em que  $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ . No entanto, tudo foi proposto para o caso particular de dois grupos. Caso as duas matrizes de covariâncias tenham eixos principais idênticos, Flury provou que a matriz  $\mathbf{A}$  é simétrica, os seus vectores próprios formam uma matriz ortogonal e  $\mathbf{\Sigma}_1$  e  $\mathbf{\Sigma}_2$  são comutáveis. A equivalência das três condições referidas anteriormente permitiu definir quatro graus de similaridade para as matrizes de covariâncias.

Foi ainda apresentada uma aplicação aos dados de Flury e Riedwyl (1983 e 1988). Estes dados resultam de um estudo sobre notas Suíças. Neles intervêm 2 grupos (notas verdadeiras e notas falsas) e 6 variáveis.

#### **Flury, B. N. (1984). Common principal components in $k$ groups.**

Neste trabalho, Flury incluiu os resultados de um relatório técnico (“technical report”) publicado em 1983, Flury (1983b). O autor utilizou a denominação de *componentes principais comuns* para o modelo que generaliza o método das *componentes principais*. A justificação apresentada para esta designação assenta no facto das colunas de  $\mathbf{\beta}$ , matriz ortogonal que diagonaliza todas as matrizes de covariâncias serem as direcções dos cosenos directores dos eixos rodados. As variáveis simultaneamente transformadas,  $\mathbf{U}_i = \mathbf{\beta}^T \mathbf{X}_i$ , foram designadas por *componentes principais comuns*. Para o caso particular de  $k$  vectores aleatórios normalmente distribuídos, Flury derivou as equações de verosimilhança e respectivas estimativas. Para a obtenção destas estimativas o autor indica a utilização de um algoritmo construído por Flury e Gautschi (1984), designado por algoritmo F-G. Derivou ainda a estatística do teste de razão de verosimilhanças para avaliar a validade do modelo das *componentes principais comuns*.

Para demonstrar o bom desempenho prático deste modelo, Flury escolheu três exemplos de dados multivariados, bastante utilizados noutros contextos da análise multivariada. Estes foram os dados Iris (Anderson, 1935 e Fisher, 1936), os referentes aos ossos de martas Norte Americanas (Jolicoeur, 1963) e as observações das notas Suíças (Flury e Riedwyl, 1983 e 1988). Flury salientou que o emprego desta

nova metodologia proporciona maior estabilidade nas estimativas (menor variância), comparativamente com o caso das *componentes principais* em separado. Isto é uma consequência directa da redução paramétrica.

Como trabalho futuro Flury dá a indicação de um modelo de CPC com restrições incorporadas que permitam a ordenação parcial das variâncias das *componentes principais*.

**Krzanowski, W. J. (1984). Principal component analysis in the presence of group structure.**

Na continuação dos trabalhos de Flury, Krzanowski propôs, o uso de outro estimador construído à custa das *componentes principais* da matriz  $\mathbf{T} = \sum_{i=1}^k \mathbf{S}_i$ , de forma a evitar o processo iterativo necessário para a estimação das matrizes de covariâncias. A justificação para o emprego deste estimador baseia-se na constatação de que, sob a validade do modelo  $H_{CPC}$ ,  $\sum_{i=1}^k \boldsymbol{\beta}^T \boldsymbol{\Sigma}_i \boldsymbol{\beta} = \sum_{i=1}^k \boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}$ , isto é,  $\boldsymbol{\beta}^T \boldsymbol{\Delta} \boldsymbol{\beta} = \boldsymbol{\Lambda}$  com  $\boldsymbol{\Delta} = \sum_{i=1}^k \boldsymbol{\Sigma}_i$ . O facto da matriz  $\mathbf{T}$  ou qualquer outra combinação linear das  $k$  matrizes ser diagonalizável com a matriz ortogonal  $\boldsymbol{\beta}$ , pode ser usado para avaliar a validade do modelo. A proposta de Krzanowski foi utilizar a matriz  $\mathbf{T}$  e a matriz de covariâncias combinada usual  $\mathbf{S}_P$ . A concordância ou não dos vectores próprios (*componentes principais*) de ambas funciona como um indicador da validade do modelo, exceptuando o caso em que o número de elementos em cada grupo é o mesmo.

O autor comparou os resultados obtidos para a matriz  $\hat{\boldsymbol{\beta}}$  com o seu processo de estimação e o preconizado por Flury (1983b). Os dados analisados em Flury (1983b) foram todos os referidos em Flury (1984), mais os de Jolicoeur e Mosimann (1960) referentes ao estudo de tartarugas.

**Flury, B. N. e Constantine G. (1985). Algorithm AS 211: the F-G diagonalization algorithm.**

Neste trabalho é apresentado o algoritmo que permite a obtenção das estimativas de máxima verosimilhança dos parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\Lambda}_i$ , para  $i = 1, \dots, k$ , referido no trabalho anterior. Com este algoritmo tenta-se encontrar uma matriz ortogonal  $\mathbf{B}$  tal que  $\Phi(\mathbf{B}) = \prod_{i=1}^k [\det(\text{diag}(\mathbf{B}^T \mathbf{S}_i \mathbf{B})) / \det(\mathbf{S}_i)]^{(n_i-1)}$  seja mínimo, com  $\mathbf{S}_i \in PDS(p)$ . O processo implementado é semelhante ao método de Jacobi, mas generalizado de forma a permitir a diagonalização de  $k$  matrizes em simultâneo. Para isso, constroem-se rotações com todos os pares de vectores ortogonais de  $\mathbf{B}$ . É apresentado o código de programa em linguagem Fortran 66.

O desempenho deste algoritmo é testado com dois exemplos, ambos com 2 populações, tendo um dimensão 2 e o outro dimensão 4.

**Flury, B. N. e Gautschi W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form.**

Neste trabalho foi exposto o algoritmo F-G com bastante pormenor. A motivação para este algoritmo baseia-se no facto da solução pretendida, que resolve o sistema (2.11), ser também a solução de um outro problema de dimensão 2.

Para construir este procedimento numérico é necessário um critério que avalie quanto as matrizes se desviam da forma diagonal. Considerando  $\mathbf{F} \in PSD(p)$ , uma medida do desvio em relação à forma diagonal é

$$\varphi(\mathbf{F}) = \frac{\det(\text{diag}(\mathbf{F}))}{\det(\mathbf{F})},$$

com  $\varphi(\mathbf{F}) \geq 1$ , sendo a igualdade alcançada quando a matriz é diagonal.

Generalizando este resultado para as  $k$  matrizes  $\mathbf{F}_i = \boldsymbol{\beta}'\mathbf{S}_i\boldsymbol{\beta} \in PSD(p)$ , uma medida dos seus desvios simultâneos em relação à forma diagonal é a apresentada em (2.13). Como referimos anteriormente, a minimização de (2.13) permite encontrar os EMV. Foi esta a via escolhida pelos autores para obter estes estimadores.

Este algoritmo é constituído por dois algoritmos separados, algoritmo F (nível externo) e algoritmo G (nível interno), que minimizam (2.13) iterativamente.

No nível F, todos os pares  $(\mathbf{f}_l, \mathbf{f}_j)$  de vectores colunas da aproximação actual de  $\mathbf{F}$  são rodados de forma a satisfazer a correspondente equação (2.11). Cada passo deste algoritmo é composto por rotações dos  $[p(p-1)/2]$  pares de colunas de  $\mathbf{F}$  a estimar.

Quanto ao nível G, neste processo iterativo procura-se encontrar uma matriz ortogonal  $(2 \times 2)$  que resolva um sistema bidimensional análogo a (2.11). Isto justifica-se pelo facto de, considerando  $\mathbf{H} = (\mathbf{f}_l : \mathbf{f}_j)$  e  $\lambda_{lj}$  como definido em (2.10),  $\mathbf{f}_l$  e  $\mathbf{f}_j$  são as soluções desejadas se e só se a matriz  $\mathbf{T}$   $(2 \times 2)$  definida por

$$\mathbf{T} = \sum_{i=1}^K (n_i - 1) \left( \frac{\lambda_{il} - \lambda_{ij}}{\lambda_{il}\lambda_{ij}} \right) \mathbf{T}_i,$$

é diagonal, com  $\mathbf{T}_i = \mathbf{H}^T \mathbf{S}_i \mathbf{H}$ .

Foi também feito um estudo sobre a convergência e unicidade do algoritmo F-G.

A eficácia deste algoritmo é testada recorrendo a um exemplo com 2 populações e 6 variáveis.

**Flury, B. N. (1986b). Asymptotic theory for common principal components analysis.**

Neste trabalho Flury desenvolveu um estudo inferencial dos estimadores de máxima verosimilhança,  $\hat{\Lambda}_i$  e  $\hat{\beta}$ , do modelo das CPC, tendo deduzido a sua distribuição assintótica. Os resultados mais relevantes encontram-se no Teorema 2.2, apresentado na Subsecção 3.2.

No que diz respeito aos valores próprios, o resultado da distribuição assintótica é facilmente obtido como consequência da convergência assintótica de  $(n_i - 1)\mathbf{S}_i$  para a normalidade, isto quando  $n_i \rightarrow \infty$ . Em contrapartida, no que refere à distribuição assintótica dos vectores próprios, a sua dedução é um pouco mais trabalhosa, mas possível com a generalização do resultado de Anderson (1963) para as *componentes principais*.

Para além dos resultados anteriores, Flury também construiu um teste de hipóteses para os vectores próprios comuns. Este novo teste generaliza o teste de Anderson (1963) o qual assume que o  $j$ -ésimo vector próprio, por exemplo, é igual a um determinado vector fixo. No teste de Flury fixa-se, em simultâneo, os valores de  $(q \leq p)$  vectores próprios.

Flury apresentou ainda um teste para avaliar a redundância de um subconjunto  $(p - q)$  das *componentes principais comuns*, assim como um outro teste para verificar a esfericidade deste subconjunto de componentes. Neste último caso pretende-se testar

$$H_s = \lambda_{i,q+1} = \dots = \lambda_{i,p}, \quad i = 1, \dots, k.$$

Por último, o autor apresentou a ilustração dos procedimentos teóricos derivados. Para isto utilizou o exemplo de Jolicoeur e Mosimann (1960), o qual é normalmente usado no contexto da análise em *componentes principais*. Para aplicar o modelo CPC, Flury considerou 2 grupos, um referente aos dados das tartarugas do género feminino e o outro com os dados do género masculino.

**Manly, B. F. J. e Rayner, J. C. W. (1987). The comparison of sample covariance matrices using likelihood ratio tests.**

Neste trabalho os autores mostram como o resultado do teste de razão de verosimilhanças, para a comparação de  $k$  matrizes de covariâncias, pode ser mais informativo adoptando uma partição especial da estatística de teste. Manly e Rayner optam por um sistema hierárquico com os seguintes níveis:

**Nível 0:** matrizes de covariâncias iguais;

**Nível 1:** matrizes de covariâncias proporcionais;

**Nível 2:** correlações iguais, i.e.,  $\mathbf{C}_i \Sigma \mathbf{C}_i$  com  $\mathbf{C}_i = \text{diag}(c_{ij})$  para  $i = 1, \dots, k$  e  $j = 1, \dots, p$ ;

**Nível 3:** matrizes de covariâncias diferentes.

Designando por  $l_i$  o máximo do logaritmo da verosimilhança sob a validade do Nível  $i$ ,  $i = 0, \dots, 3$ , e como os quatro níveis constituem uma estrutura encaixada, vem que  $l_0 \leq l_1 \leq l_2 \leq l_3$ . Além disso, caso o modelo do Nível  $i$  seja correcto,  $T_{i+1} = 2(l_{i+1} - l_i)$  tem distribuição assintótica qui-quadrado com o número de graus de liberdade igual à diferença entre os parâmetros a estimar sob os modelos  $(i+1)$  e  $i$ . A estatística soma,  $T^* = T_1 + T_2 + T_3$ , é a usualmente utilizada nos testes de razão de verosimilhanças para avaliar as diferenças entre matrizes de covariâncias.

Os autores propõem um teste sequencial, começando por testar a validade de  $T_3$ . Quando o resultado do teste é significativamente grande assume-se que as matrizes de covariâncias têm correlações distintas, terminando o teste. Caso o resultado do teste não seja significativo para rejeitar esta hipótese, então  $T_2$  é testado. Quando o resultado de  $T_2$  é significativo, assume-se que as variâncias são diferentes mas as correlações iguais terminado o teste. No caso contrário, segue-se com o teste  $T_1$ . Um resultado significativo neste teste leva a concluir a proporcionalidade das matrizes de covariâncias, e no caso contrário é assumido a igualdade das  $k$  matrizes.

Foi feita uma aplicação a um exemplo com 2 grupos em dimensão 8.

**Flury, B. N. (1987a). Two generalizations of the common principal components model.**

Foi neste trabalho que pela primeira vez Flury propôs o modelo das *componentes principais comuns parciais*. Este modelo, o Nível 4 do sistema hierárquico de Flury apresentado na Secção 2, assume que apenas  $q$  vectores próprios são comuns a todas as matrizes  $\Sigma_i$ , e permite que os restantes  $(p-q)$  difiram com o grupo. Esta ideia tem algumas semelhanças com as de Krzanowski (1979). Como vimos anteriormente, este último autor, tentou analisar o subespaço gerado pelos primeiros  $q$  vectores próprios de cada grupo.

Foi feita a dedução dos EMV para o modelo CPCP. O estudo da validade deste modelo foi feito com base em duas estatísticas de razão de verosimilhanças, sendo uma exacta e outra aproximada. Além disso, foi proposto ainda um teste para avaliar a validade de um modelo análogo ao preconizado por Krzanowski (1979), designado por modelo com  $q$  *componentes principais com espaço gerado comum*.

Para aplicação dos resultados obtidos, o autor seleccionou os dados estudados por Airoldi e Hoffmann (1984). Neste caso tem-se um problema com 4 grupos em di-

mensão 4.

**Flury, B. N. (1987b). A hierarchy of relationships between covariance matrices.**

O autor introduziu pela primeira vez o seu sistema hierárquico, formado com os 5 níveis apresentados na Secção 2, deste capítulo. Além destes modelos, é ainda referido o modelo das  $q$  *componentes principais com espaço gerado comum*, Flury (1987a), representado por  $CS(q)$  e localizado entre o Nível 4 e o Nível 5. Para esta

MODELO A	MODELO B	g.l.
igualdade	proporcionalidade	$k - 1$
proporcionalidade	CPC	$(p - 1)(k - 1)$
CPC	CPCP( $q$ )	$\frac{1}{2} (k - 1) (p - q) (p - q - 1)$
CPCP( $q$ )	CS( $q$ )	$\frac{1}{2} (k - 1) q (q - 1)$
CS( $q$ )	arbitrárias	$(k - 1) q (q - 1)$

**Tabela 2.4:** Decomposição do teste de razão de verosimilhanças para a igualdade de  $k$  matrizes de covariâncias.

estrutura encaixada, o autor propôs novos testes de razão de verosimilhanças. Na mesma perspectiva de Manly e Rayner (1987), Flury indica uma decomposição da estatística de teste para avaliar a validade, ou não, dos níveis por ele sugeridos. Na Tabela 2.4 apresentamos um esquema com os testes propostos, assim como os respectivos graus de liberdade (g.l.) associados a cada um deles.

Para avaliar o desempenho do novo sistema hierárquico, Flury utilizou os dados de Flury (1987a).

**Keramidas, E. M., Devlin, S. J. e Gnanadesikan, R. (1987). A graphical procedure for comparing the principal components of several covariance matrices.**

Segundo estes autores, até à data deste trabalho pouca atenção foi dispensada ao estudo do grau de semelhança entre as *componentes principais* de vários grupos. De forma a contrariar esta tendência, foi proposto um método gráfico para comparar várias matrizes de covariâncias, em termos da orientação dos seus vectores próprios. O processo gráfico indicado aplica a técnica designada por *papel de probabilidade* para a distribuição Gama. Com este método é possível quantificar o valor do ângulo entre dois vectores. Um destes vectores é uma das *componentes principais* e o outro é um vector chamado “típico”. Para Keramidas *et al.* (1987), um vector “típico” é aquele que mais se aproxima da mesma *componente principal* em todos os  $k$  grupos.



Para a construção deste vector “típico” os autores sugerem procurar, para cada grupo, um vector que minimize o ângulo entre si e o  $j$ -ésimo vector próprio.

Os autores propõem ainda uma extensão deste método, para a comparação de um conjunto de vários vectores próprios em simultâneo nos  $k$  grupos, i.e., quando a estrutura  $p$ -dimensional inicial permite a decomposição em vários subespaços esféricos.

No entanto, este método é puramente analítico, no sentido em que não foi desenvolvido nenhum modelo estatístico. Além disso, tem a desvantagem de só poder ser utilizado quando os valores próprios estão bastantes separados.

Com o objectivo de avaliar o comportamento do processo gráfico proposto, os autores desenvolveram estudos de Monte Carlo para várias situações. Variaram as matrizes de covariâncias, o número de observações, o número de populações, a dimensionalidade do espaço das variáveis e a distribuição dos dados. Utilizaram ainda dois exemplos reais. Um deles trata da avaliação feita por alunos universitários aos seus professores, com 7 grupos e dimensão 10. O outro exemplo com 15 grupos em dimensionalidade 5 foi também estudado por Chen *et al.* (1974).

**Airoidi, J. P. e Flury, B. N. (1988). An application of common principal component analysis to cranial morphometry of *Microtus californicus* and *M. ochrogaster* (Mammalia, Rodentia).**

Para estes autores, os modelos das CPC e das CPCP são boas alternativas para modelar o grau de semelhanças entre as *componentes principais* de vários grupos. Como já vimos, estes métodos admitem que todas as componentes, ou parte delas, são comuns aos  $k$  grupos, sendo as discrepâncias, essencialmente, consequência da variabilidade amostral.

Embora vários outros autores tenham também tratado exemplos com  $k$  grupos, normalmente as *componentes principais* são construídas a partir da matriz de covariâncias combinada ou então agrupam-se todos os dados num só grupo. Neste trabalho é bem realçado que estas vias não conduzem a bons resultados.

Foram apresentados os resultados teóricos que permitem o cálculo dos erros padrão associados aos valores e vectores próprios.

Os autores realçaram ainda que, quando as *componentes principais* variam com o grupo então, as CPC vão sofrer maior influência dos grupos com valores próprios mais separados.

No exemplo escolhido, tridimensional e com 4 grupos de pequenos roedores (géneros

feminino e masculino das espécies “*Microtus californicus*” e “*Microtus ochrogaster*”), foram comparadas as análises em CP, CPC e CPCP. Dos resultados obtidos, foi possível concluir que os novos métodos originam estimativas das *componentes principais* com menores erros padrões do que a análise em CP para cada grupo isoladamente.

**Schott, J. R. (1988). Common principal components subspace in two groups.**

Este trabalho desenvolveu-se tendo em linha de conta que a análise em *componentes principais* é um bom instrumento para diminuir a dimensão do problema em questão. Esta ideia pode ser facilmente alargada para problemas com 2 ou mais grupos. Nestes casos, o subespaço gerado por um subconjunto das  $p$  *componentes principais*,  $q$ , deve ser o mesmo em todos os grupos.

Este problema tinha sido estudado por Flury (1987b). O que Schott indica neste trabalho é um procedimento aproximado para testar a hipótese de espaço comum, para problemas com 2 grupos.

Assim, sendo

$$\Sigma_1 = \mathbf{P}_* \Lambda_1 \mathbf{P}_*^T, \quad \Sigma_2 = \mathbf{Q}_* \Lambda_2 \mathbf{Q}_*^T,$$

e

$$\mathbf{P}_* = (\mathbf{P}_{*1} \mathbf{P}_{*2}), \quad \mathbf{Q}_* = (\mathbf{Q}_{*1} \mathbf{Q}_{*2}),$$

com  $\mathbf{P}_{*1}$  e  $\mathbf{Q}_{*1}$  matrizes  $(p \times q)$ , a hipótese a testar é

$$H_0(q): \mathbf{P}_{*1} = \mathbf{Q}_{*1} \mathbf{R}_1, \text{ com } \mathbf{R}_1 \text{ } (q \times q) \text{ e ortogonal.}$$

Seja ainda  $\mathbf{K}$  uma matriz ortogonal de tal modo que

$$((n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2)/(n_1 + n_2 - 2) = \mathbf{K} \Delta \mathbf{K},$$

com  $\Delta$  diagonal,  $\mathbf{K} = (\mathbf{K}_1 \mathbf{K}_2)$  e  $\mathbf{K}_1$  e  $\mathbf{K}_2$  definidas como anteriormente. No caso de  $H_0(q)$  ser verdadeiro vem que,  $\mathbf{K}_1 \mathbf{K}_1^T = \mathbf{P}_{*1} \mathbf{P}_{*1}^T = \mathbf{Q}_{*1} \mathbf{Q}_{*1}^T$ , e o subespaço comum pode ser calculado como  $\mathbf{V} = \{\mathbf{u} \in \mathbb{R}^p : \mathbf{K}_1 \mathbf{K}_1^T \mathbf{u} = \mathbf{u}\}$ . Sendo  $T$  a estatística deste teste, o autor usou uma aproximação para  $T$ .

Convém realçar que o procedimento preconizado por Schott não é igual ao proposto por Flury (1987b). O teste  $H_{CS}(q)$  de Flury (1987b) não especifica quais são as *componentes principais* que formam o subespaço para cada grupo.

Schott fez ainda um estudo de simulação e uma aplicação a um exemplo real. Esse exemplo foi o estudado por Flury (1987a).

**Flury, B. N. (1988). Common principal components and related multivariate models.**

Neste livro encontra-se toda a teoria desenvolvida por Flury no estudo de estruturas de covariâncias. Todos os resultados teóricos, distribuições assintóticas e testes de razão de verossimilhanças, para o sistema hierárquico deste autor, são apresentados. Incluem-se também todos os exemplos referidos nos trabalhos prévios deste autor. O algoritmo F-G é apresentado, assim como todo o estudo de convergência deste algoritmo.

No Capítulo 7 o autor apresenta novos problemas e algumas questões em aberto. O primeiro dos tópicos sugeridos por Flury para trabalho futuro, trata da estimação robusta das CPC. Isto demonstra a relevância da estimação robusta das *componentes principais comuns*. O autor realça bem o quanto é importante encontrar novos procedimentos estatísticos, robustos, que não sejam muito afectados pela presença de observações discordantes ou pela não normalidade dos dados.

**Krzanowski, W. J. (1990). Between-groups analysis with heterogeneous covariance matrices: The common principal components model.**

Neste trabalho Krzanowski aplicou o modelo das *componentes principais comuns*, como ponto de partida para a análise de *variáveis canónicas*. Como é sabido, nas *variáveis canónicas* é imposta a igualdade das matrizes de covariâncias. Admitindo que  $p$  variáveis são recolhidas em  $n$  indivíduos, e “a priori” são feitas divisões dos indivíduos em  $k$  grupos, o objectivo da análise em *variáveis canónicas* é analisar as diferenças entre grupos. Normalmente, este método estatístico assume que  $\Sigma_i = \Sigma$ , para todo o  $i$ . Mas, muitas vezes na prática esta situação não se verifica. Como alternativa, Krzanowski sugere a utilização dos modelos intermédios do sistema hierárquico de Flury (1987b, 1988, págs. 60–62), em particular o modelo das CPC. Sob a validade do modelo CPC, o autor desenvolveu dois métodos para analisar as diferenças entre os grupos. Um deles é uma extensão do método de Campbell e Atchley (1981), o qual usa uma análise sequencial em *componentes principais*. O outro método baseia-se numa nova definição de distância entre grupos, em vez da usual distância de Mahalanobis, trata-se da distância de Matusita (1956).

O autor utilizou os dados referidos em Krzanowski (1979) para ilustrar estes novos procedimentos.

**Schott, J. R. (1991). Some tests for common principal components subspaces in several groups.**

O autor generalizou o teste que tinha proposto em 1988, só com 2 grupos, para  $k$

grupos. Para isso, utilizou novamente uma estatística de teste aproximada.

Como é do conhecimento geral em alguns problemas é conveniente construir as *componentes principais* a partir das matrizes de correlações. Schott, neste trabalho, construiu também um teste de hipóteses que se aplica nesta situação.

O autor apresentou ainda os resultados de um estudo de simulação que permitiram avaliar a qualidade da aproximação da sua estatística de teste.

**Klingenberg, C. P. e Zimmermann, M. (1992). Static, ontogenetic, and evolutionary allometry: A multivariate comparison in nine species of water striders.**

Neste trabalho foi aplicado o método das CPC a dados biológicos. O objectivo deste estudo é a análise de dados morfológicos recolhidos em 5 larvas de 9 espécies dos insectos “Heteroptera: Gerridae”. Além disso, ainda foram incluídas 2 espécies criadas em laboratório. Foram formados 3 grupos:

1. **“static allometry”**: variação da estrutura dentro da larva;
2. **“ontogenetic allometry”**: variação da estrutura dentro de determinada espécie, incluindo todas as larvas;
3. **“evolutionary allometry”**: variação da estrutura entre espécies, dentro de determinada larva.

Devido à forma como o problema está estruturado, os dados não podem seguir a distribuição normal multivariada. O que deve modelar estas observações é uma mistura de várias distribuições. Os autores referem que, por exemplo, os dados da Classe 3 formam 5 agrupamentos distintos para cada espécie em estudo (Zimmermann, 1987). Assim, Klingenberg e Zimmermann realçam que não é correcto aplicar a metodologia de Flury para calcular os erros padrões ou efectuar o teste de razão de verosimilhanças para avaliar o modelo CPC. Por esse motivo aplicaram métodos de “jackknife” e de “bootstrap” (Efron e Tibshirani, 1986).

Os autores salientam um problema que pode ocorrer com o método “bootstrap”, consequência das reduzidas dimensões amostrais. Nestas situações, existe uma grande probabilidade de pelo menos uma das matrizes de covariâncias amostrais não ser definida positiva (Daudin *et al.*, 1988).

Klingenberg e Zimmermann tentaram explicar eventuais relações entre os 3 grupos construídos através das *componentes principais comuns*.

**Yvan, K. H. e Bentler, P. M. (1994). Test of linear trend in eigenvalues of**

***k* covariance matrices with applications in common principal components analysis.**

Como referimos anteriormente, Flury (1986b) derivou uma estatística de teste para avaliar a esfericidade do subconjunto das  $(q - p)$  *componentes principais comuns*. Neste teste avalia-se a igualdade dos valores próprios correspondentes, para cada grupo. Yvan e Bentler, por seu lado, propõem um novo teste. Na opinião destes autores é mais realista os valores próprios apresentarem uma tendência linear, do que a igualdade sugerida por Flury. Neste caso existem  $a_i$ ,  $b_i$  e  $(x_1^{(i)}, \dots, x_{i_t}^{(i)})$ , tais que para  $(j \leq i_t)$  a hipótese a testar é traduzida por

$$H_l : \lambda_{i,q+j} = a_i + b_i x_j^{(i)}, \quad i = 1, \dots, k.$$

Caso esta hipótese não seja rejeitada, os autores sugerem o teste  $H_{b_i} : b_i = 0$ . Note-se que a não rejeição desta última hipótese leva-nos à proposta de Flury (1986b).

Com a validade de  $H_l$  a decomposição espectral de  $\Sigma_i$  fica

$$\Sigma_i = \sum_{j=1}^q \lambda_{ij} \beta_j \beta_j^T + \sum_{t=1}^{p-q} \left( a_i + b_i x_t^{(i)} \right) \beta_{q+t} \beta_{q+t}^T, \quad i = 1, \dots, k,$$

condição esta que os autores introduziram na equação referida em (2.8). Criaram assim um novo sistema de equações de verosimilhança. Efectuaram o estudo dos EMV resultantes deste modelo, o modelo das CPC com a hipótese  $H_l$  adicional. Foi feito ainda um estudo das propriedades assintóticas dos estimadores, assim como construída a estatística para testar  $H_l$ .

Yvan e Bentler também desenvolveram uma metodologia semelhante para o caso da validade das CPCP.

Para exemplificação da metodologia desenvolvida, Yvan e Bentler escolheram o exemplo de Jöreskog e Sörbom (1979, pág. 202), o qual tinha sido tratado por Jöreskog mas no contexto da *análise factorial*. Seleccionaram 3 dos 4 grupos com dimensão 9 e concluíram que o seu modelo é adequado.

**Klingenberg, C. P., Neuenschwander, B. E. e Flury, B. N. (1996). Ontogeny and individual variations: Analysis of patterned covariance matrices with common principal components.**

Os autores propuseram uma generalização do modelo CPC a dados longitudinais, o que poderá ser designado por análise em *componentes principais comuns* para grupos dependentes. Como Klingenberg *et al.* (1996) realçaram, em muitos estudos de crescimento recolhem-se medições das mesmas espécies em diferentes estados

ontogenéticos. Consequentemente, esses estados não podem formar grupos independentes. Nestes casos, existe um conjunto de  $k$  medidas (por exemplo o grupo ou o estado de crescimento), cada um com as mesmas  $p$  variáveis, para todas as observações.

Consideram-se então  $kp$  variáveis simultâneas, com matriz de covariâncias ( $kp \times kp$ ) com  $(k \times k)$  blocos. A dimensão de cada um dos blocos é  $(p \times p)$ . No caso de  $k = 2$ , vem

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Esta generalização do modelo CPC assume que todas as matrizes  $\Sigma_{ij}$  são diagonalizáveis com a mesma transformação ortogonal. No caso de  $k = 2$  a matriz de covariâncias ( $2p \times 2p$ ) depois de transformada fica

$$\begin{bmatrix} \beta^T \Sigma_{11} \beta & \beta^T \Sigma_{12} \beta^T \\ \beta^T \Sigma_{21} \beta & \beta^T \Sigma_{22} \beta \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},$$

com todas as matrizes  $\Lambda_{ij}$  diagonais.

Foi feita uma aplicação prática desta metodologia a um exemplo da área da biologia. Os dados são relativos aos insectos “Heteroptera: Gerridae”, insectos que habitam à tona da água, em 6 estados diferentes de desenvolvimento (grupos). Foram observadas 4 variáveis em cada um dos estados e concluíram que o modelo proposto é vantajoso na análise deste tipo de dados.

**Steppan, S. J. (1997). Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal components analysis.**

Algumas técnicas quantitativas permitem compreender a macroevolução dos organismos. Normalmente, estas técnicas assumem que as variâncias e covariâncias genéticas se mantêm constantes. Para avaliar a veracidade desta suposição, o autor introduziu duas novas técnicas filogenéticas. Estas novas técnicas, para o estudo da evolução da estrutura de covariâncias, têm naturezas distintas. Uma delas é heurística enquanto que a outra segue uma abordagem estatística.

Steppan desenvolveu o seu estudo baseado em 28 populações de roedores “Phyllotis”. Nestas populações intervêm 13 subespécies (diagnosticáveis) de 6 espécies biológicas. Foram formadas classes de acordo com o desgaste e a erupção molar dos “Phyllotis”. Este estudo foi desenvolvido com base na observação de 24 medições do crânio e do maxilar destes roedores. O autor refere também a taxinomia que utilizou.

Steppan construiu matrizes de correlações com todas as variáveis. Comparou todos

os possíveis pares de matrizes de correlações para o seu esquema filogenético. Além disso, efectuou também uma análise em *componentes principais comuns* com 13 variáveis. Os resultados destas duas análises não foram concordantes. Contudo, o autor refere que as CPC deve ser encaradas como uma alternativa viável para conjuntos de dados com grandes estruturas filogenéticas.

**Schott, J. R. (1998). Estimating correlation matrices that have common eigenvectors.**

Neste artigo foi apresentado um método que permite obter estimadores de máxima verosimilhança das matrizes de correlações de  $k$  grupos, quando todas as matrizes de correlações têm os mesmos vectores próprios. O autor está assim a propor uma variante do modelo das *componentes principais comuns*. Esta alternativa é importante para problemas onde a escala (ou ordem de grandeza) das variáveis em estudo é muito diferente.

Considerando  $\mathbf{\Omega}_i$  a matriz de correlações de população  $i$ , então

$$H_{0r} : \mathbf{\Omega}_i = \mathbf{P}\mathbf{D}_i\mathbf{P}^T ,$$

é o modelo preconizado por Schott. De forma análoga ao modelo de Flury, apresentado em (2.3),  $\mathbf{P}$  é a matriz ortogonal construída com os vectores próprios comuns e  $\mathbf{D}_i$  são matrizes diagonais com os valores próprios de cada grupo.

Schott, sugere aplicar o algoritmo F-G para encontrar uma estimativa inicial de  $\mathbf{P}$  e  $\mathbf{D}_i$ . Além disso, indicou um processo para obter estimadores mais refinados destes parâmetros. Construiu ainda um teste assintótico, de Wald, para avaliar a validade de  $H_{0r}$ .

A adequação do modelo foi avaliada com dois exemplos. Os primeiros dados são os clássicos dados Iris, e o segundo conjunto de dados pode encontrar-se em Morrison (1990) ou em Altman (1968). No último conjunto de dados intervêm 2 grupos com dimensão 5.

**Sengupta, S. e Boyle, J. S. (1998). Using common principal components for comparing GCM simulations.**

Neste trabalho foi apenas feita uma aplicação do modelo das *componentes principais comuns*, i.e., não foram adicionadas contribuições teóricas sobre este tema. A necessidade do uso das CPC foi sentida num trabalho da área da meteorologia. Foi necessário confrontar os resultados da investigação de vários AMIP (“Atmospheric Model Intercomparison Project”). Do AMIP resultam infra-estruturas para a comparação de modelos atmosféricos circulares gerais (GCM). Existe um número elevado

de modelos para comparar e, como Sengupta e Boyle referem, a técnica das CPC é bastante útil.

No estudo desenvolvido por estes autores intervieram 30 modelos AMIP mais algumas observações da precipitação nos Estados Unidos. No total foi necessária a comparação de 31 matrizes de covariâncias com dimensão 120, porque foi recolhida a média da precipitação em 120 meses.

Os autores concluíram que o modelo das *componentes principais comuns* é bastante eficaz para a comparação dos resultados de um elevado número de GCM.

**Phillips, P. C. e Arnold, S. J. (1999). Hierarchical comparison of generic variance-covariance matrices. I. Using the Flury hierarchy.**

Este é mais um trabalho que evidencia a importância prática do modelo das *componentes principais comuns*. Neste caso a área de aplicação foi a Biologia.

Os autores pretenderam efectuar um estudo sobre matrizes de covariâncias genéticas, designadas por matrizes-G. Como referimos anteriormente, o estudo destas matrizes é uma via para o conhecimento da evolução das espécies. Uma possível abordagem é aplicar o sistema hierárquico de Flury (1987b, 1988, págs. 60–62). No entanto, Phillips e Arnold salientam a necessidade da normalidade das observações para os teste de hipóteses de Flury (1987b, 1988, págs. 82–85). Além disso, o número de graus de liberdade da estatística de razão de verosimilhanças, sob a hipótese nula, tem de ser conhecido. Para dados do tipo genético normalmente estas duas condições não são satisfeitas. Em geral, a forma da distribuição é desconhecida. Por outro lado, o número dos graus de liberdade está relacionado como número de famílias em estudo, e o número de indivíduos de cada família. A distribuição da estatística de teste parece ser um qui-quadrado mas, o número de graus de liberdade inadequado. Para resolver este problema os autores propõem um novo teste de hipóteses construído com técnicas de reamostragem. É desenvolvido um procedimento aleatório que redistribui aleatoriamente as famílias na população.

Como aplicação prática, Phillips e Arnold escolheram dados referentes ao estudo da “*Thamnophis elegans*” (cobras americanas). O objectivo é a comparação das matrizes-G de duas populações que se encontram separadas aproximadamente 300 Km, e com sistemas ecológicos diferentes. Uma das populações habita no interior, no estado da Califórnia, enquanto a outra é do litoral. Foram observadas 6 variáveis.

**Arnold, S. J. e Phillips, P. C. (1999). Hierarchical comparison of generic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*.**



Este trabalho segue a mesma linha do anterior, ou seja, aplica a metodologia hierárquica de Flury (1987b, 1988, págs. 60–62). O objectivo dos autores é a comparações das duas populações de cobras “*Thamnophis elegans*” referidas em Phillips e Arnold (1999). É simplesmente feita uma aplicação prática do modelo das CPC.

**Achermann, R. R. e Cheverud, J. M. (2000). Phenotypic covariance structure in *Tomarins* (*genus saguinus*): A comparison of variation patterns using matrix correlation and common principal components analysis.**

Achermann e Cheverud (2000) seguiram o trabalho de Steppan (1997). Os métodos usados neste trabalho foram as *componentes principais comuns* e as matrizes de correlações. Pretende-se relacionar espécies de saguins da família “*Callitrichidae*”, género “*Saguinus*”. Neste estudo intervêm saguins adultos e são feitas 39 medições em 848 crânios. São analisadas 6 espécies diferentes.

O modelo das CPC foi rejeitado, i.e., as diferentes espécies de saguins variam mas o padrão de variação não se encaixa em nenhum dos modelos de Flury (1987b, 1988, págs. 60–62).

Todavia, Achermann e Cheverud (2000) alertam para o facto de ambos as análises, correlações e CPC, serem muito úteis nos estudos morfológicos.

Com tudo o que foi exposto ao longo deste capítulo, em especial a última subsecção, pensamos que está bem evidente a importância da técnica estatística das *componentes principais comuns*.

# Capítulo 3

## Estimação Robusta

### 3.1 Introdução

Com este capítulo pretendemos introduzir, embora sem grande detalhe, algumas noções da estatística robusta. Estes conceitos foram fundamentais e permitiram construir os novos procedimentos estatísticos que apresentamos nos Capítulos 4, 5 e 6.

Uma possível definição do termo robusto pode ser encontrado no livro “A dictionary of statistical terms” (de Kendall e Buckland, 1971) segundo o qual

*“... a statistical procedure is described as robust if it is not very sensitive to departures from the assumptions on which it depends”.*

Muitos dos modelos estatísticos são construídos supondo a normalidade, a independência e a idêntica distribuição das variáveis (vectores) aleatórias(os) em estudo, mas, regra geral, estas suposições são apenas aproximações da realidade. Infelizmente, as consequências destas aproximações podem ser drásticas afectando o desempenho dos estimadores clássicos usados nos procedimentos inferenciais.

Já em 1960 no trabalho de Tukey intitulado “A survey of sampling from contaminated distributions” este evidenciou o facto dos estimadores que são bastante eficientes assintoticamente no cenário da normalidade, como  $\bar{X}$ , perderem esta característica ao deslocarmo-nos, mesmo que ligeiramente, para uma vizinhança dessa distribuição. Daí a necessidade dos procedimentos robustos. Esta constatação é de longa data. No século XIX alguns estatísticos aplicados, como foi o caso de Bessel (1818) e um pouco mais tarde de Newcomb (1886, 1895), sentiram que a

falha de algumas suposições em modelos paramétricos conduzia a maus resultados inferenciais. Contudo, o termo robusto só foi introduzido na estatística na década de 50 do século XX, no trabalho de Box (1953). Desde essa data, vários estatísticos têm desenvolvido esforços na construção de procedimentos estatísticos robustos em diversas áreas da estatística.

A quantificação do grau de robustez dos novos procedimentos é possível devido à teoria de robustez que teve início com Huber (1964) e Hampel (1968) e foi apresentada em pormenor nos livros de Huber (1981) e de Hampel *et al.* (1986). Nestes livros encontram-se os conceitos de robustez que têm servido de referência essencial aos estatísticos que trabalham nesta área. Destes novos conceitos talvez os mais utilizados sejam a *função de influência*, uma noção local ou infinitesimal de robustez, e o *ponto de ruptura*, uma noção global de robustez. O primeiro deles, a *função de influência*, vai ter um papel extremamente importante no nosso trabalho, pois os estudos de robustez desenvolvidos apoiaram-se neste conceito.

As noções de *função de influência* e de *ponto de ruptura* foram introduzidos na tese de doutoramento de Hampel (1968) e posteriormente apresentadas nas publicações de Hampel (1971, 1974). No entanto, de *ponto de ruptura* já existia uma versão anterior da autoria de Hodges (1967).

Uma vez que uma das principais razões que conduzem à utilização de procedimentos robustos é a presença de possíveis observações discordantes (“outliers”) apresentamos na Secção 3 um breve desenvolvimento deste tópico.

Além disso, também apresentamos os estimadores robustos de dispersão que utilizamos/referimos neste trabalho, quer para o caso univariado (Secção 6), quer para o caso multivariado (Secção 8).

## 3.2 Algumas generalidades

Para o desenvolvimento das técnicas estatísticas consideram-se usualmente amostras da população a estudar de dimensão  $n$ ,  $(X_1, \dots, X_n)$ . Supõe-se também que as variáveis aleatórias (v.a.) são independentes e identicamente distribuídas com  $X$ , podendo tomar valores num espaço amostra  $\Omega$ . Estas são distribuídas de acordo com a “verdadeira” distribuição de probabilidade  $F$ . A distribuição  $F$  pode, por exemplo, pertencer à família de distribuições paramétricas  $\{F_\theta, \theta \in \Theta\}$ , com  $\theta$  desconhecido e pertencente ao espaço parâmetro  $\Theta$ . Um dos objectivos da estatística é a estimação de  $\theta$ .

No entanto, não existe geralmente uma verdadeira distribuição de probabilidade que traduza o comportamento da(s) variável(eis) em estudo na população. Devemos assim considerar um conjunto mais alargado de distribuições para representar a distribuição da(s) característica(s) em estudo.

Em alternativa à estatística clássica, a estatística robusta considera que a população em estudo segue uma distribuição  $G$  pertencente a uma vizinhança  $\varepsilon$  (vizinhança- $\varepsilon$ ) de  $F$  (modelo paramétrico assumido).

Uma forma para se definir uma vizinhança- $\varepsilon$ , com  $0 < \varepsilon < 1$ , é considerar como Huber (1964) sugeriu, seguindo o trabalho de Tukey (1960),

$$GE(F, \varepsilon) = \{G : G = (1 - \varepsilon)F + \varepsilon H, H \in \mathcal{F}\} , \quad (3.1)$$

conhecido como “gross error model” e onde  $\mathcal{F}$  é um espaço geral de distribuições sobre  $\Omega$ .

Aqui, o facto do modelo  $F$  não se verificar exactamente é evidenciado pela contaminação  $H$  com peso  $\varepsilon$ . O resultado é uma mistura de duas distribuições. Ao gerarmos uma amostra de dimensão  $n$  a partir desta mistura, esperamos que em média  $(1 - \varepsilon) \times n$  das observações sejam oriundas de  $F$ , enquanto que as restantes ( $\varepsilon \times n$ ), as ditas observações contaminantes, provenham da distribuição  $H$ . No entanto, o “gross error model”,  $GE(F, \varepsilon)$ , não traduz todas as alternativas realísticas do modelo  $F$ . Este apenas inclui as alterações consequência dos erros grosseiros. Como Hampel (1968) referiu, este modelo em sentido heurístico inclui apenas “metade” da “vizinhança total”. Como alternativa, Hampel propôs considerar vizinhanças em topologias “fracas”. Para as construir podemos pensar nas topologias que admitem métricas em  $\mathcal{F}$ . Por exemplo, podemos definir em  $\mathcal{F}$  uma vizinhança usual, usando uma bola aberta (ou fechada),

$$B_\varepsilon(F) = \{G : d(G, F) < \varepsilon\} , \quad (3.2)$$

com  $\varepsilon > 0$ ,  $F$  e  $G \in \mathcal{F}$ , e  $d(., .)$  uma métrica em  $\mathcal{F}$ .

Uma das métricas mais usadas é a de Prohorov (1956). Esta foi utilizada posteriormente por Hampel (1971) para avaliar a robustez dos estimadores, tendo sido também usada na teoria das probabilidades em teoremas limite. A distância de Prohorov entre  $F$  e  $G$  é definida por

$$d_{Pr}(F, G) = \inf\{\varepsilon > 0 : G(A) \leq F(A^\varepsilon) + \varepsilon, \forall A \in \mathcal{B}\} , \quad (3.3)$$

onde  $\mathcal{B}$  representa a  $\sigma$ -álgebra gerada pelos abertos de  $\Omega$  e

$$A^\varepsilon = \{x \in \Omega : \inf_{y \in A} d(x, y) \leq \varepsilon\}. \quad (3.4)$$

Com esta métrica Hampel (1971) classifica um estimador de  $\theta$  de robusto quando os pequenos desvios em relação a  $F$  se traduzem em pequenas alterações distribucionais, como veremos na Subsecção 5.1.

Todavia, a situação do “gross error model” é muitas vezes preferível segundo Hampel *et al.* (1986, págs. 400-401). O comportamento dos estimadores no  $GE(F, \varepsilon)$ , ou pelo menos alguns deles como é o caso dos estimadores de Huber (1964), é bastante similar ao observado em vizinhanças mais realistas. Além disso, já que a definição de vizinhança é bastante arbitrária e como o “gross error model” tem a vantagem de simplificar a análise, do ponto de vista matemático, parece justificada a sua utilização. Foi este o contexto usado em muitos dos trabalhos de Huber e de Hampel, e actualmente continua a ser a base de muitos estudos robustos. A teoria da robustez assim desenvolvida (“a teoria estatística dos modelos paramétricos aproximados”) considera-se um ramo da estatística paramétrica (mais pormenores encontram-se em Hampel, 2000).

Convém realçar que embora de alguma forma próximas a estatística robusta não é um ramo da estatística não paramétrica. Esta última inclui métodos de distribuição livre, como por exemplo o teste dos sinais, que são supostamente correctos para uma larga classe de distribuições. No entanto, esta não é a situação da estatística robusta. Ao contrário do caso não paramétrico, a estimação robusta desenrola-se com o suporte de um determinado modelo paramétrico, que aproxima a realidade. É neste cenário que se pretendem estimar os parâmetros associados ao modelo. No entanto, por vezes existe uma certa confusão entre estas duas situações, como referem Portnoy e He (2000). Isto acontece porque muitas vezes os procedimentos não paramétricos são para certos modelos procedimentos robustos. Este é o caso da mediana que apesar de ser uma estatística não paramétrica é bastante robusta para a estimação do centro de um modelo paramétrico simétrico.

Fundamental para a estimação robusta de parâmetros é o conceito de funcional. Um funcional não é mais que uma aplicação (medida descritiva,  $T(F)$ ) definida num espaço convexo de distribuições. Usualmente o espaço amostra ( $\Omega$ ) é o conjunto  $\mathcal{R}$  mas pode também ser um conjunto de categorias ou um espaço Euclidiano de dimensão superior. O domínio normalmente inclui todas as distribuições empíricas e também a hipotética verdadeira distribuição. Os funcionais estatísticos são uma abstracção conveniente, pois algumas propriedades estatísticas podem ser expressas em termos das suas propriedades analíticas.

Para se formalizar este conceito vamos considerar uma amostra aleatória de dimensão  $n$ ,  $(X_1, \dots, X_n)$  de variáveis aleatórias identicamente distribuídas de acordo com

$G$ . Seja  $G_n = \frac{1}{n} \sum_{i=1}^n \Delta_{x_i}$ , onde  $\Delta_x$  indica a distribuição degenerada em  $x$ , a função de distribuição empírica da amostra. Um estimador natural da função de distribuição  $G$  é frequentemente  $G_n$ . Como estimadores de um parâmetro  $\theta$  consideram-se as estatísticas  $T_n = T_n(X_1, \dots, X_n)$ . Portanto, para cada amostra de dimensão  $n$ , um estimador não é mais do que uma sequência das estatísticas  $\{T_n; n \geq 1\}$ . A maior parte dos resultados estatísticos podem ser expressos em termos de  $G_n$  e  $n$ . No entanto, seria bastante vantajoso podermos usar uma função que fosse independente da dimensão amostral,  $n$ . Assim, sempre que é possível escrever para uma dada amostra a igualdade

$$T_n(X_1 \dots X_n) = T(G_n) , \quad (3.5)$$

dizemos que o estimador  $T_n$  é equivalente ao funcional  $T$ . Uma vez que  $G_n$  pode ser visto como uma medida de probabilidade (aleatória), podemos considerar frequentemente os estimadores como (ou substituí-los por) funcionais, pelo menos aproximadamente ou assintoticamente. Neste último caso admitimos que existe um funcional  $T$ , aplicação do espaço de todas as distribuições de probabilidade  $\mathcal{F}(\Omega)$  com uma métrica geral, para  $\mathbb{R}$  (ou  $\mathbb{R}^p$  no caso de  $\theta$  ser um vector de dimensão  $p > 1$ ), que sob certas condições de regularidade (Huber 1964, pág. 76) verifica a convergência fraca

$$\lim_{n \rightarrow \infty} T(G_n) = T(G) \quad (\text{em probabilidade}). \quad (3.6)$$

O comportamento de  $T$  na vizinhança de  $F$  pode ser avaliado através do comportamento de  $T(G_n)$  pois a função de distribuição empírica ( $G_n$ ) tem forte probabilidade de se encontrar nessa vizinhança. Está assim justificado este procedimento.

Como exemplos de funcionais podemos considerar:

## 1. Média

Neste caso considera-se  $T_1(G_n) = \int x dG_n(x) = \sum_{i=1}^n \frac{X_i}{n} = \bar{X}$ .

Note-se que a segunda igualdade é uma consequência de  $G_n$  atribuir probabilidades  $\frac{1}{n}$  a cada observação  $X_i$ .

## 2. Mediana

Seja  $G^{-1}$  a função inversa definida da forma usual, ou seja  $G$  com  $G^{-1}(p) = \inf \{x : G(x) \geq p, 0 < p < 1\}$ . Considere-se  $p = \frac{1}{2}$  e um funcional  $T_2$  para o qual  $T_2(G) = G^{-1}(\frac{1}{2})$ .

*Da aplicação do funcional à função de distribuição empírica resulta*

$$T_2(G_n) = G_n^{-1} \left( \frac{1}{2} \right) = X_{(\lfloor \frac{n+1}{2} \rfloor)},$$

*sendo  $\lfloor \cdot \rfloor$  a característica do número.*

Para a média o estimador e o funcional são equivalentes, para a mediana esta equivalência verifica-se apenas para  $n$  ímpar. Existem situações onde nem esta equivalência não se verifica. Normalmente, isto acontece como consequência da discretização das observações, ou devido a algum termo de correcção do enviesamento da ordem de  $1/n$ . Vejamos o próximo exemplo.

### 3. Variância

*Vamos considerar o estimador centrado da variância com a média desconhecida  $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$ . Não é possível encontrar um funcional que aplicado à função de distribuição empírica iguale o estimador. No entanto, considerando o funcional*

$$T_3(G_n) = \int (x - T_1(G_n))^2 dG_n(x),$$

*verificamos que*

$$\lim_{n \rightarrow \infty} S^2 = T_3(G), \quad (\text{em probabilidade}).$$

*Mas, para amostras de dimensão finita, resulta*

$$T_3(G_n) = \frac{n-1}{n} S^2.$$

Devido ao facto do funcional  $T$  estar definido no espaço  $\mathcal{F}$  é necessário avaliar o seu comportamento em  $G$  (pertencente a uma vizinhança do modelo paramétrico  $F$ ), pois é aí que hipoteticamente se distribui(em) a(as) característica(s) em estudo. Além disso, pode perfeitamente acontecer que o “verdadeiro” modelo não se encontre em  $G$ . Assim, é necessário exigirmos outro tipo de convergência de forma a que o estimador  $T_n$  (ou o funcional  $T$ ) estime correctamente o parâmetro, sob o modelo paramétrico. Dizemos então que o funcional é consistente segundo Fisher ou Fisher-consistente (Kallianpur e Rao, 1955) para o parâmetro  $\theta$  se e só se,

$$T(F_\theta) = \theta, \quad \forall \theta \in \Theta. \quad (3.7)$$

Vejamos os exemplos que se seguem.

## 1. Média

Considere-se a situação em que  $F_\theta$  representa a distribuição exponencial com valor esperado  $\theta$ . O funcional média ( $T_1$ ) é Fisher-consistente pois

$$T_1(F_\theta) = \theta, \quad \forall \theta \in \mathbb{R}^+.$$

## 2. Mediana

Considere-se agora para a mesma distribuição o funcional mediana ( $T_2$ ). Este por sua vez já não é Fisher-consistente (para  $\theta$ ) pois

$$T_2(F_\theta) = F^{-1}\left(\frac{1}{2}\right) = \theta \ln 2.$$

Um funcional Fisher-consistente para  $\theta$ , baseado na mediana, é

$$T(F_\theta) = \frac{T_2(F_\theta)}{\ln 2}.$$

Por último, é também possível provar que, sob certas condições de regularidade, se verifica a seguinte convergência

$$n^{1/2} [T(G_n) - T(G)] \xrightarrow{d} N(0, V), \quad (3.8)$$

i.e., a sequência  $n^{1/2} [T(G_n) - T(G)]$  tem uma distribuição assintoticamente normal com média assintótica 0 e  $V = \text{ASVAR}(T, G)$ . Note-se que o último parâmetro denominado talvez abusivamente de variância assintótica não é

$$\lim_{n \rightarrow \infty} \text{VAR}(T(G_n)),$$

mas

$$\lim_{n \rightarrow \infty} n \text{VAR}(T(G_n)).$$

Então para  $n$  elevado (finito) verifica-se  $\text{VAR}(T(G_n)) \approx \text{ASVAR}(T, G)/n$ .

## 3.3 Observações discordantes e erros grosseiros

Um problema que qualquer estatístico prático provavelmente defronta é a presença de erros grosseiros (“gross errors”) e/ou observações discordantes (“outliers”). Os “gross errors” são todos os erros consequência de alguma anomalia. Por exemplo, falhas no equipamento, erros de transmissão, erros de leitura ou escrita, mistura de



duas condições experimentais ou falsa categorização no delineamento experimental, poderão ser a origem deste tipo de erros. A frequência dos “gross errors” depende obviamente da forma como os dados foram recolhidos. No entanto, em Hampel *et al.* (1986, págs. 25–28) refere-se que nos dados de ciências exactas a frequência destes erros é normalmente de 1 a 10% enquanto que para as ciências não exactas é usual observar-se um valor superior a 10%.

Por outro lado, por “outlier” entende-se qualquer observação que não siga o padrão dominante da maioria dos dados, daí a designação de observação discordante. No entanto, esta não é uma noção que se possa considerar rigorosa ou bem definida uma vez que é difícil dizer a partir de que valor uma dada observação está discordante com a maioria. Este conceito é extremamente importante e cada vez mais surgem estudos com a finalidade de encontrar processos que conduzam à identificação de “outliers” quer para o caso univariado quer para o caso multivariado. Esta última situação é mais complexa, uma vez que no caso de se trabalhar com dados em dimensão  $p \geq 2$  um “outlier” pode não ser “outlier” em qualquer das coordenadas univariadas. Daí a maior dificuldade na sua identificação.

Este problema é ainda abordado neste trabalho no Capítulo 6, onde são propostas duas medidas para a identificação de “outliers” (ou observações influentes na estimação das CPC). Estas novas propostas também podem ser utilizadas para  $k = 1$ , portanto também permitem a detecção de observações “outliers”/influentes nas *componentes principais*.

Os “outliers” não são o mesmo que os “gross errors”. É fácil fazer-se a distinção entre este dois conceitos embora na prática possa surgir alguma confusão. Muitos dos “gross errors” mostram-se como “outliers”, mas alguns podem estar escondidos entre as observações que se consideram “limpas”.

Quanto aos “outliers” muitos deles são “gross errors” mas alguns podem ser observações correctas, embora estejam discordantes com a maioria. Os “outliers” podem também ser consequência de uma distribuição com caudas longas ou até serem as observações mais relevantes. Como exemplo, podemos referir a descoberta do buraco de ozono na Antárctida onde essas observações foram claramente identificadas como “outliers”, (Hampel, 2000). Daí não ser aconselhável retirar (rejeitar) os “outliers” sem uma segunda análise dos dados, pois até os próprios “gross errors” podem conter alguma informação útil.

Existem dois objectivos principais quando se analisam os dados dando relevância aos “outliers”; um deles tenta compreender o padrão descrito pela maioria dos dados com a acomodação dos “outliers” e o outro preocupa-se com a identificação de todos os

“outliers”, sendo os “outliers” mais relevantes objecto de um estudo pormenorizado (Hampel, 1991). O primeiro destes dois objectivos é precisamente uma das preocupações da estatística robusta. Por outro lado, os resíduos de um ajustamento robusto são uma base muito boa para a identificação de pontos especiais (Hampel, 2000).

### 3.4 A abordagem minimax de Huber

Foi Huber (1964) quem deu o primeiro passo na teoria de robustez. Segundo ele

*“A convenient measure of robustness for asymptotically normal estimators seems to be the supremum of the asymptotic variance when  $G$  (the distribution of the sample) ranges over some suitable set of the underlying distributions”.*

Introduziu uma nova classe de estimadores, estimadores-M, que generalizam os estimadores da máxima verosimilhança (EMV).

Os estimadores-M, do parâmetro  $\theta$  do modelo de localização  $F_\theta(x) = F(x - \theta)$ , definem-se por

$$T_n = (X_1, \dots, X_n) \quad \text{que minimiza} \quad \sum_{i=1}^n \rho(x_i - T_n), \quad (3.9)$$

onde  $\rho$  é alguma função regular não constante. Considerando  $\rho' = \psi$  então o estimador  $T_n$  também satisfaz a equação implícita,

$$\sum_{i=1}^n \psi(x_i - T_n) = 0. \quad (3.10)$$

Huber mostrou que sob certas condições de regularidade o estimador  $T_n = (X_1, \dots, X_n)$  converge quase certamente ( $n \rightarrow \infty$ ) para um funcional  $T$  verificando

$$\int \psi(x - T(G)) dG(x) = 0. \quad (3.11)$$

Provou também que a distribuição de  $n^{1/2}(T_n(X_1, \dots, X_n) - T(G))$  é assintoticamente normal com média e variância assintótica, respectivamente, 0 e  $ASVAR(T, G) = \frac{\int \psi^2(x - T(G)) dG(x)}{(\lambda'(T(G)))^2}$ . No denominador da  $V(T, G)$  consideramos  $\lambda(y) = \int \psi(x - y) dG(x)$ ,  $\forall y \in \mathbb{R}$  (diferenciável em  $y = T(G)$ ) e com  $\lambda'(T(G)) < 0$ .

A escolha da função  $\psi$  é preponderante e pode tornar, ou não, o estimador robusto. Para escolhas convenientes da função  $\psi$  estes estimadores têm a vantagem de se mostrarem bastante eficientes sob a validade de certos modelos paramétricos. Além disso, quando nos deslocamos para uma vizinhança próxima desses modelos paramétricos os efeitos ao nível distribucional são pequenos. Dada a importância da função  $\psi$ , normalmente a variância assintótica dos estimadores-M é denotada por  $ASVAR(\psi, G)$  em vez de  $ASVAR(T, G)$ .

Vejamos os seguintes exemplos de estimadores-M.

### 1. EMV

Se  $f$  representar a função densidade de probabilidade (f.d.p.) os EMV correspondem a considerar  $\rho_{MV} = -\log f$  e  $\psi_{MV} = -\frac{f'}{f}$  e atingem o mais pequeno valor possível para  $ASVAR(\psi_{MV}, F)$ , i.e., igualam o inverso da medida de informação de Fisher. Quando  $F = \Phi$  (função de distribuição da normal standardizada) a média aritmética será a solução encontrada considerando  $\psi_{MV}(x) = x$ , resultando  $ASVAR(\psi_{MV}, \Phi) = 1$ .

### 2. Mediana

A mediana será a solução resultante quando  $\rho_{med}(x) = |x|$  e  $\psi_{med}(x) = \text{sign}(x)$ . Agora para  $F = \Phi$  temos  $ASVAR(\psi_{med}, \Phi) = \frac{\pi}{2}$ .

Note-se que para uma distribuição  $F$  geral com  $F^{-1}(\frac{1}{2}) = 0$  resulta que  $ASVAR(\psi_{med}, F) = \frac{1}{(2f(0))^2}$ .

### 3. Estimador de Huber (abordagem minimax)

O estimador de Huber foi descoberto como a solução do problema minimax de encontrar o estimador que minimiza o máximo da variância assintótica para a distribuição normal standard,  $\Phi$ . Huber (1964) considerou o modelo  $GE(\Phi, \varepsilon)$ , i.e., o “gross error model” no caso particular de  $F = \Phi$  e  $H$  uma distribuição simétrica em torno de 0. Definiu um estimador com  $\psi_{Hu(k)} = \min(k, \max(x, -k))$ ;  $k \in \mathbb{R}^+$ . Considerando  $\Psi$  o conjunto com as funções  $\psi$  regulares, quando  $\varepsilon$  e  $k$  estão relacionados por  $\frac{\varepsilon}{1-\varepsilon} = \frac{2}{k}\phi(k) - 2\Phi(-k)$  com  $\phi = \Phi'(x)$ , resulta então

$$ASVAR(\psi_{Hu(k)}, F_{Hu}) = \min_{\psi \in \Psi} \max_{G \in GE(\Phi, \varepsilon)} ASVAR(\psi, G). \quad (3.12)$$

Na expressão anterior,  $F_{Hu}$  representa a distribuição que Huber designou por distribuição “menos” favorável. Esta distribuição é o elemento de  $GE(F, \varepsilon)$  que torna tão difícil quanto possível a estimação do centro

da distribuição, i.e., maximiza a variância assintótica do estimador de localização do parâmetro  $\theta$  para a pior contaminação. Portanto, é para a distribuição  $F_{Hu}$  que se atinge o menor valor da medida de informação de Fisher,  $I(\theta) = \int (f'/f)^2 f dx$ . Esta é uma distribuição Gaussiana no centro e exponencial nas caudas.

Como resultado o estimador de Huber é determinado com

$$\psi_{Hu(k)}(x) = \begin{cases} x, & |x| < k \\ k \operatorname{sign}(x), & |x| \geq k \end{cases}. \quad (3.13)$$

Este estimador intuitivamente pode ser interpretado como a média de uma amostra modificada (transformada pela função  $\psi$ ) onde os “outliers” são integrados no “miolo” dos dados de uma forma suave.

Huber (1964) também resolveu o problema minimax para o estimador de escala e para o estimador simultâneo de localização e escala.

Algumas das críticas mais frequentes à teoria minimax de Huber são o facto desta se desenvolver no “gross error model”, com a agravante de exigir uma mistura simétrica, e o critério de optimização ser baseado numa quantidade assintótica. Além de se poder argumentar que a teoria minimax é demasiado pessimista. Para mais referências sobre eventuais objecções a esta abordagem de Huber ver Hampel *et al.* (1986, págs. 397–403).

Posteriormente, Huber (1965, 1968 e Huber e Strassen, 1973), desenvolveu uma segunda teoria de robustez considerando vizinhanças mais gerais do modelo normal. Esta teoria tem, como já se referiu, por um lado, a desvantagem do difícil manuseamento matemático e, por outro, a vantagem de não ser construída com a imposição do “dogma” do modelo simétrico.

### 3.5 A abordagem infinitesimal de Hampel

Hampel (1968, 1971, 1974) propôs uma outra metodologia robusta apoiada em três novos conceitos:

1. continuidade do funcional  $T$  (robustez qualitativa);
2. curva de influência ou função de influência (robustez infinitesimal);
3. ponto de ruptura (robustez quantitativa ou robustez global).

Nas subsecções seguintes estes temas serão objecto de um estudo mais aprofundado. Será dado um maior destaque à *função de influência*, pois a abordagem seguida nesta tese baseia-se essencialmente neste conceito de robustez.

### 3.5.1 Robustez qualitativa

A *robustez qualitativa* está bastante relacionada com a continuidade da estatística quando esta é vista como um funcional num espaço com uma topologia “fraca”. Temos a seguinte definição.

**Definição 3.1 (robustez qualitativa, Hampel, 1971)** Dizemos que a sequência de estimadores  $\{T_n : n \geq 1\}$  possui *robustez qualitativa* em  $T$  se para todo  $\varepsilon > 0$  existir  $\delta > 0$  tal que

$$d_{Pr}(F, G) < \delta \Rightarrow d_{Pr}(\mathcal{D}_F(T_n), \mathcal{D}_G(T_n)) < \varepsilon, \quad (3.14)$$

onde  $\mathcal{D}_F(T_n)$  representa a distribuição da estatística sob a validade de  $F$  e  $\mathcal{D}_G(T_n)$  sob a validade de  $G$ .

Os dois teoremas seguintes relacionam este conceito de robustez com continuidade.

**Teorema 3.1** A sequência de estimadores  $\{T_n : n \geq 1\}$  que é contínua em  $F$  e para a qual todo o  $T_n$  é uma função contínua das observações goza de *robustez qualitativa*.

**Teorema 3.2** Seja  $T$  um funcional equivalente aos estimadores  $\{T_n : n \geq 1\}$ .  $T$  é contínuo (com respeito à distância de Prohorov) para toda a distribuição  $F$  se e só se  $\{T_n : n \geq 1\}$  é *qualitativamente robusta* para toda a distribuição  $F$  e ainda satisfaz  $d_{Pr}(\mathcal{D}_F(T_n), F) \xrightarrow{n \rightarrow \infty} 0$ .

Vejamos a aplicação deste conceitos nos dois funcionais seguintes.

#### 1. Média

*Considere-se o funcional média restrito à distribuição  $G$  de forma a resultar numa quantidade finita, i.e.,  $T(G) = \int x dG(x)$  é finito. Apesar desta imposição este funcional não é contínuo e portanto também não goza de robustez qualitativa. Para verificar este facto, tome-se  $GE(F, \varepsilon)$  no caso particular de  $H = \Delta_x$  (medida de probabilidade que atribui massa*

1 no ponto  $x$ ). Facilmente se observa que  $D_{Pr}(F, GE(F, \varepsilon)) \leq \varepsilon \forall x$  mas  $|T(F) - T(GE(F, \varepsilon))| = \varepsilon |x - T(F)|$ , quantidade esta que cresce proporcionalmente com o crescimento de  $x$  podendo tornar-se arbitrariamente grande.

## 2. Mediana

Considere-se o funcional mediana  $T(G) = G^{-1}(\frac{1}{2})$  o qual está definido para todo o  $G$ . Este funcional é contínuo em  $F$  se e só se  $\{x : F(x) \geq \frac{1}{2}\}$  contém apenas um ponto. Nesse caso o funcional mediana é robusto (qualitativamente) em  $F$ .

### 3.5.2 Função de influência

A *função de influência* (IF) não é mais do que uma ferramenta heurística que permite avaliar o comportamento local de um funcional  $T$ . Para se dar uma ideia desta função vamos considerar uma amostra concreta  $(x_1, \dots, x_{n-1})$  com função de distribuição empírica representada por  $F_{n-1}$ . Suponha-se que se acrescentou uma nova observação  $x$ . O que se pretende é encontrar uma forma de avaliar as alterações sofridas localmente no funcional  $T(F_{n-1})$  consequência desta nova observação. Como  $F_n = ((n-1)F_{n-1} + \Delta_x)/n$  podemos pensar em usar a *função de sensibilidade amostral*, também conhecida por *curva de sensibilidade amostral*,  $SC_n$  (Tukey, 1970–71) dada por  $n(T(F_n) - T(F_{n-1}))$ . Assim vem que

$$SC(x) = \left[ T \left( \left( 1 - \frac{1}{n} \right) F_{n-1} + \frac{1}{n} \Delta_x \right) - T(F_{n-1}) \right] \bigg/ \frac{1}{n}. \quad (3.15)$$

Considerando que  $F_{n-1}$  pode substituir  $F$  (teorema de Glivenko-Cantelli, ver por exemplo Billingsley, 1986, pág. 275) e no caso do funcional ser “bem” comportado, a função anterior avalia a alteração sofrida em  $T$  consequência do novo ponto em  $x$  com massa  $\varepsilon = 1/n$ , isto numa amostra grande da distribuição  $F$ . Com a passagem ao limite resulta a definição de *função de influência*.

**Definição 3.2 (função de influência, Hampel, 1968, 1974)** A *função de influência* (IF) do funcional  $T$  em  $F$  é dada por

$$IF(x, T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)F + \varepsilon \Delta_x) - T(F)}{\varepsilon}, \quad (3.16)$$

nos pontos  $x \in \Omega$  para os quais este limite existe.

Uma justificação mais formal desta função baseia-se na ideia de diferenciação de um funcional que remonta a von Mises (1947). Considere-se  $G$  uma distribuição “próxima” de  $F$ , com respeito à distância de Prohorov, e ambas no mesmo espaço de distribuições. Sob certas condições de regularidade é válido o desenvolvimento de von Mises de primeira ordem de  $T$  em  $F$ , dado por

$$T(G) = T(F) + \int a_1(x) dG(x) + R(G - F), \quad (3.17)$$

onde  $a_1$  é o chamado núcleo da primeira derivada de von Mises e  $R(G - F)$  representa o resto.

No caso particular de  $G$  ser  $F_{x,\varepsilon} = (1 - \varepsilon)F + \varepsilon\Delta_x$  e considerando apenas os dois primeiros termos do desenvolvimento (3.17) resulta (notando que  $\int a_1(u)dF(u) = 0$ )

$$\begin{aligned} T(F_{x,\varepsilon}) &\cong T(F) + \int a_1(u)F_{u,\varepsilon}(u)du = \\ &= T(F) + \varepsilon \int a_1(u)\Delta_x(u)du = T(F) + \varepsilon a_1(x). \end{aligned} \quad (3.18)$$

Resolvendo esta equação em ordem a  $a_1(x)$  e fazendo  $\varepsilon \rightarrow 0$ , obtém-se a definição da *função de influência* atrás referida, uma vez que

$$a_1(x) = \lim_{\varepsilon \rightarrow 0} \left[ \frac{T(F_{x,\varepsilon}) - T(F)}{\varepsilon} \right] = \text{IF}(x, T, F).$$

Tem-se então o seguinte desenvolvimento linear de primeira ordem

$$T(G) \cong T(F) + \int \text{IF}(x, T, F)dG(x). \quad (3.19)$$

Com a expressão anterior avaliada em  $F = G$  resulta o valor esperado nulo de  $\text{IF}(x, T, F)$  podendo  $dG(x)$  ser substituído por  $d(G - F)(x)$ .

Existe uma relação importante entre a *função de influência* e a variância assimp-tótica do funcional. Esta pode ser encontrada considerando uma amostra aleatória  $\mathbf{X} = (X_1, \dots, X_n)$  com função de distribuição empírica  $F_n$ . Se substituirmos em (3.17)  $G$  por  $F_n$  vem

$$\begin{aligned} T(F_n) &\cong T(F) + \int \text{IF}(x, T, F)dF_n(x) + R(F_n - F) = \\ &= T(F) + \frac{1}{n} \sum_{i=1}^n \text{IF}(x_i, T, F) + R(F_n - F). \end{aligned} \quad (3.20)$$

No caso de  $\sqrt{n}R(F_n - F) \xrightarrow{p} 0$  resulta pelo teorema do limite central e lema de Slutsky (Ferniholz, 1983, págs 40 e 41) que  $[n^{1/2}(T(F_n) - T(F))]$  tem distribuição assimp-tótica normal com média zero e variância assimp-tótica (ASVAR) dada por

$$\text{ASVAR}(T, F) = \frac{1}{n} \text{VAR} \left( \sum_{i=1}^n \text{IF}(X_i, T, F) \right) = E_F [\text{IF}(X, T, F)^2]. \quad (3.21)$$

Quando o funcional é consistente segundo Fisher é ainda possível mostrar que

$$E [\text{IF} (X, T, F_\theta)^2] \geq \frac{1}{I(\theta)}, \quad (3.22)$$

onde  $I(\theta)$  representa a medida de informação de Fisher.

A condição necessária e suficiente para que se atinja a igualdade na expressão (3.22) é que  $\text{IF}(x, T, F)$  seja proporcional a  $\frac{\partial}{\partial \theta} \log f(x)$ , onde  $f(x)$  representa a f.d.p..

Tomando apenas o termo de primeira ordem em (3.20) resulta a seguinte aproximação entre uma estimativa de  $\theta$ ,  $\hat{\theta}$ , e o parâmetro  $\theta$

$$\hat{\theta} \cong \theta + \frac{1}{n} \sum_{i=1}^n \text{IF}(x_i, T, F), \quad (3.23)$$

sendo fácil observar que o factor  $\text{IF}(x_i, T, F)$  é um indicador da influência que o valor  $x_i$  tem no resultado da estimativa  $\hat{\theta}$ . Foi este facto que levou Hampel (1968, 1974) a apelidar a função IF de *curva de influência*. Inicialmente, Hampel utilizou *curva de influência* em vez de *função de influência* mas, posteriormente, em Hampel *et al.* (1986, pág 84) foi utilizada a designação de função visando a generalização para dimensões superiores.

Numa perspectiva matemática a *função de influência* é o conjunto de todas as derivadas parciais de  $T$  em  $F$  na direcção de  $(\Delta_x - F)$ . Isto é consequência directa da derivada de uma função poder ser avaliada à custa das derivadas parciais, caso exista derivada. Reportando-nos ao funcional  $T$ , definido num espaço de dimensão infinita, a derivada de  $T$ , quando existe, pode então ser descrita através do conjunto (infinito) das derivadas parciais em misturas da forma  $F_{x,\varepsilon}$ .

Para os funcionais existem vários conceitos de diferenciabilidade, como é o caso da de Gâteaux, Hadamard ou Fréchet. O último destes conceitos é de todos o mais forte não sendo por isso muito utilizado. Todavia a diferenciabilidade de Fréchet é um conceito bastante útil pois é uma possível via para a dedução da distribuição limite dos funcionais. A diferenciabilidade segundo Gâteaux é, destas três, a que exige condições mais fracas, contudo para existir a *função de influência* estas condições ainda podem ser “relaxadas”. Assim, para a generalidade dos funcionais a *função de influência* pode ser calculada sem preocupação com as condições de regularidade.

A *função de influência* apresenta-se assim como uma ferramenta essencial na área da estatística e pode ser calculada para a maioria dos funcionais mais utilizados. Tem ainda a vantagem de poder ser utilizada na dedução da variância assintótica dos funcionais, e de permitir a construção de estimadores robustos com propriedades es-



pecíficas. Para Huber (1972) esta é mesmo a ferramenta heurística mais importante para obter estimadores com “boas” propriedades de robustez.

É importante tentar construir funcionais com *função de influência* limitada, pois caso contrário os funcionais não gozam de *robustez qualitativa*. Com este procedimento tenta-se limitar a influência das “más” observações nos funcionais. Ao mesmo tempo, será de todo conveniente que a *função de influência* também seja aproximadamente proporcional a  $\frac{\partial}{\partial \theta} \log f(x)$  fazendo com que o estimador relacionado com o funcional não perca demasiada eficiência.

Vejamos de seguida a *função de influência* de alguns funcionais.

## 1. Média

Considere-se o funcional média  $T_1(F) = \int x dF(x) = \mu(F)$ . Temos que

$$T_1(F_{x,\varepsilon}) = (1 - \varepsilon) \mu(F) + \varepsilon x,$$

e portanto a *função de influência* vem

$$IF(x, \mu(F), F) = \lim_{\varepsilon \rightarrow 0} \frac{(1 - \varepsilon) \mu(F) + \varepsilon x - \mu(F)}{\varepsilon} = x - \mu.$$

## 2. Variância

Seja  $T_2(F) = \int (x - T_1(F))^2 dF(x) = \sigma^2(F)$  o funcional variância.

Como

$$T_2(F_{x,\varepsilon}) = (1 - \varepsilon) \sigma^2(F) + \varepsilon (x - \mu(F))^2,$$

resulta a seguinte expressão para a *função de influência*

$$\begin{aligned} IF(x, \sigma^2(F), F) &= \lim_{\varepsilon \rightarrow 0} \frac{(1 - \varepsilon) \sigma^2(F) + \varepsilon (x - \mu(F))^2 - \sigma^2(F)}{\varepsilon} = \\ &= (x - \mu)^2 - \sigma^2. \end{aligned}$$

Como se pode verificar nas duas situações anteriores, as *funções de influência* resultantes não dependem da distribuição paramétrica considerada (desde que  $\mu(F)$  e  $\sigma^2(F)$  existam). Contudo, estes dois funcionais apresentam a grande desvantagem de terem *funções de influência* que tendem para infinito com  $x$ . Portanto, ambos os funcionais, assim como os estimadores equivalentes, não apresentam características robustas, embora para o caso do modelo normal sejam os mais eficientes.

Observe-se agora o caso de dois funcionais onde a *função de influência* é limitada.

### 3. Mediana

Seja  $T_3(F) = F^{-1}\left(\frac{1}{2}\right)$  o funcional mediana e assumamos que  $F$  tem f.d.p.  $f$  contínua e positiva em  $F^{-1}\left(\frac{1}{2}\right) = \text{med}$ . Para a construção da função de influência é necessário avaliar  $T(F_{x,\varepsilon}) = F_{x,\varepsilon}^{-1}\left(\frac{1}{2}\right) = h(\varepsilon)$ .

É fácil verificar que (ver por exemplo Staudte e Sheather, 1990, pág. 56)

$$F_{x,\varepsilon}^{-1}(u) = \begin{cases} F^{-1}\left(\frac{u}{1-\varepsilon}\right) & , \quad u < (1-\varepsilon)F(x) \\ x & , \quad (1-\varepsilon)F(x) \leq u < (1-\varepsilon)F(x) + \varepsilon \\ F^{-1}\left(\frac{u-\varepsilon}{1-\varepsilon}\right) & , \quad (1-\varepsilon)F(x) + \varepsilon \leq u \end{cases}$$

Neste caso podemos calcular a função de influência através da  $\left[\frac{\partial}{\partial \varepsilon}h(\varepsilon)\right]_{\varepsilon=0}$ .

Considerando o caso particular de  $u = \frac{1}{2}$  e para  $(1-\varepsilon)F(x) + \varepsilon \leq \frac{1}{2}$  vem,

$$\frac{\partial}{\partial \varepsilon}h(\varepsilon) = \frac{\partial}{\partial \varepsilon}F^{-1}\left[\left(\frac{1}{2} - \varepsilon\right)/(1-\varepsilon)\right] = \frac{(-\frac{1}{2})/(1-\varepsilon)^2}{f\{F^{-1}[(\frac{1}{2} - \varepsilon)/(1-\varepsilon)]\}}.$$

Assim, este ramo da função de influência, para  $x < \text{med}$ , será

$$IF(x, \text{med}, F) = \frac{(-\frac{1}{2})/(1-\varepsilon)^2}{f\{F^{-1}[(\frac{1}{2} - \varepsilon)/(1-\varepsilon)]\}} \Big|_{\varepsilon=0} = \frac{-1}{2f(\text{med})}.$$

Procedendo de forma análoga para as restantes situações resulta,

$$IF(x, \text{med}, F) = \begin{cases} \frac{-1}{2f(\text{med})} & , \quad x < \text{med} \\ \frac{1}{2f(\text{med})} & , \quad x > \text{med} \\ 0 & , \quad x = \text{med} \end{cases}$$

Como se pode constatar, qualquer contaminação no ponto  $x = \text{med}$  é inconsequente para o desempenho do funcional mediana. Por outro lado, contaminando à esquerda da mediana observa-se uma alteração constante e com efeito negativo. Contrariamente, para valores de  $x$  superiores à mediana a alteração fixa resultante tem efeito positivo. O valor destes efeitos é, como se vê na expressão anterior, inversamente proporcional ao valor de  $f(\text{med})$ . Por outro lado, a existência de uma descontinuidade em  $x = \text{med}$  indica uma elevada sensibilidade local e pode ser uma desvantagem deste funcional.

Contrariamente aos exemplos anteriores, este funcional tem características de robustez, pois o seu valor não é fortemente alterado com a contaminação numa observação.

#### 4. Estimador–M de localização

Considere-se o funcional  $T(F)$  equivalente ao estimador definido pela equação implícita (3.10), com  $\int \psi(x - T(F))dF(x) = 0$ .

A função de influência deste funcional pode ser construída substituindo na expressão anterior  $F$  por  $F_{x,\varepsilon}$  e posteriormente avaliando a sua derivada em ordem a  $\varepsilon$  no ponto zero. Assim, temos

$$\frac{\partial}{\partial \varepsilon} \left[ \int \psi(x - T(F))d(F + \varepsilon(\Delta_x - F)) \right] = 0.$$

Quando o núcleo deste integral é limitado e mensurável resulta que

$$\frac{\partial}{\partial \varepsilon} \int \psi(x - T(F))dF(x) + \int \psi(x - T(F))d(\Delta_x - F) = 0.$$

Aplicando a regra da derivação da função composta e avaliando o seu resultado com  $\varepsilon = 0$ , tem-se

$$\int \frac{\partial}{\partial \theta} [\psi(x - T(F))] \big|_{\theta=T(F)} dF(x) \frac{\partial}{\partial \varepsilon} [T(F_{x,\varepsilon})]_{\varepsilon=0} + \psi(x - T(F)) = 0.$$

Assim vem que

$$IF(x, T, F) = \frac{\psi(x - T(F))}{\int \psi'(x - T(F))dF(x)},$$

se o denominador for diferente de zero e sendo  $\psi'$  a derivada de  $\psi$ .

Como o denominador é uma constante, podemos interpretar o numerador como contendo a informação essencial sobre a forma da função de influência. Assim, observa-se que a função de influência é proporcional à função  $\psi$  que define a equação de estimação. Esta função pode então ser escolhida de forma a limitar a influência das observações discordantes e, adicionalmente, pode também controlar a eficiência do estimador.

Em conclusão, com o estudo de IF é possível construir estimadores, pelo menos para alguns modelos particulares, bastante eficientes e com “boas” propriedades de robustez.

### 3.5.3 Medidas de robustez derivadas da função de influência

Com o objectivo de estudar as propriedades de robustez dos estimadores equivalentes aos funcionais, Hampel (1968, 1974) introduziu ainda três importantes quantidades

sumárias calculadas a partir da *função de influência*. A primeira e talvez a mais importante é a *sensibilidade a grandes erros* (“gross-error sensitivity”), usualmente denotada por  $\gamma^*$ . Esta é o supremo do valor absoluto de IF e intuitivamente mede, aproximadamente, a pior influência possível que a contaminação poderá ter no valor assintótico do estimador.

A *sensibilidade a grandes erros* pode também ser encarada como um limite superior para o enviesamento do estimador com contaminações pontuais. Assim, é bastante importante atingir-se, para um dado estimador, um valor finito de  $\gamma^*$ . Portanto, quando a *sensibilidade a grandes erros* é ilimitada,  $\gamma^* = \infty$ , o estimador é completamente intolerante aos “outliers”. Um simples “outlier” arruina mesmo o desempenho do estimador. Daí a seguinte definição:

**Definição 3.3 (estimador B-robusto, Rousseeuw, 1981)** O estimador equivalente ao funcional  $T$  é *B-robusto* em  $F$  se  $\gamma^* = \gamma^*(T, F) < \infty$ .

A segunda medida sumária está relacionada com pequenas flutuações nas observações. A medida do efeito local dessas alterações, devidas por exemplo a arredondamentos ou agrupamentos, é dada pela mais pequena constante de Lipschitz à qual IF obedece. Esta constante designada por *sensibilidade local* (“local-shift-sensitivity”) é dada por

$$\lambda^* = \sup_{x \neq y} \frac{|\text{IF}(y, T, F) - \text{IF}(x, T, F)|}{|y - x|}.$$

Note-se que contrariamente à *sensibilidade a grandes erros* agora um valor infinito de  $\lambda^*$  poderá ter uma consequência ligeira no estimador. Isto poderá ser uma consequência da standardização, ou da descontinuidade da *função de influência*. A mediana, por exemplo, tem  $\lambda^* = \infty$  e é um estimador robusto, o mais *B-robusto* no caso do modelo normal.

Por último, é ainda importante sabermos quando um estimador rejeita os “outliers” e a que distância isso acontece. Em linguagem da *função de influência* isto significa que deverá existir uma região onde fora da qual a influência da contaminação é nula. No caso da distribuição ser simétrica em torno de zero resulta que esta constante, designada por *ponto de rejeição*, é igual a

$$\rho^* = \inf \{r > 0 : \text{IF}(x, T, F) = 0, \text{ se } |x| > r\}.$$

Caso não exista um  $r$  para o qual isto acontece, por definição, considera-se  $\rho^* = \infty$ .

### 3.5.4 Versões empíricas da função de influência

A *função de influência*, como se viu anteriormente, é baseada em funcionais que podem coincidir com os estimadores só assintoticamente. No entanto, existem também algumas versões amostrais da *função de influência*, das quais iremos descrever três.

Uma das versões amostral foi proposta por Hampel *et al.* (1986, págs. 92–95) e designada por *função de influência empírica* (EIF). Para a construção desta função existem duas alternativas, uma com a adição e outra com a substituição de uma observação.

Dado um determinado estimador  $T_n$ , uma amostra  $x_1, \dots, x_{n-1}$ , e uma observação adicional  $x$ , a *função de influência empírica* é

$$\text{EIF}(x) = T_n(x_1, \dots, x_{n-1}, x). \quad (3.24)$$

A versão construída com a substituição, funciona a partir de uma amostra de dimensão  $n$  e onde, por exemplo, a observação  $x_n$  é substituída por uma observação arbitrária  $x$ . De seguida define-se  $\text{EIF}_n$  como em (3.24). Esta versão é particularmente importante quando a observação  $x_n$  é um “outlier”.

Tukey (1970–71) por outro lado, popularizou a *curva de sensibilidade* como uma ferramenta para avaliar consequências sofridas no estimador aquando da alteração de uma observação numa amostra finita. Esta pode também ser criada com a substituição ou a adição de uma observação. No último caso considerando o estimador  $T_n$  e a amostra  $x_1, \dots, x_{n-1}$ , define-se a *curva de sensibilidade* como uma função da observação adicional  $x$  através de

$$\text{SC}(x) = n[T_n(x_1, \dots, x_{n-1}, x) - T_{n-1}(x_1, \dots, x_{n-1})].$$

Note-se que quando o estimador é um funcional, i.e.,  $T_n(x_1, \dots, x_n) = T(F_n)$  para todo o  $n$ , resulta a definição apresentada em (3.15).

Por último, a terceira alternativa é o *jackknifing* (Quenouille, 1956; Tukey, 1958). Considere-se um estimador  $T_n = T_n(x_1, \dots, x_n)$  equivalente a um funcional, ou seja  $T_n(x_1, \dots, x_n) = T(F_n)$ .

Por definição, o  $i$ -ésimo *pseudo-valor jackknifed*  $T_{ni}^*$  é

$$T_{ni}^* = nT_n(x_1, \dots, x_n) - (n-1)T_{n-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Inicialmente Quenouille (1956) propôs este método com a finalidade de possibilitar uma redução no enviesamento da estimação. Este autor observou que um estimador

$T_n$ , do parâmetro  $\theta$ , pode frequentemente ver o seu enviesamento expandido em termos da dimensão amostral por

$$E(T_n - \theta) = a_1 n^{-1} + a_2 n^{-2} + o(n^{-3}).$$

Quando isto acontece, Quenouille propôs utilizar-se o estimador  $T_n^*$  construído fazendo a média da amostra das *pseudo-observações*  $(T_{n1}^*, \dots, T_{nn}^*)$ . Normalmente, o estimador  $T_n^*$  consegue ter menor enviesamento que o estimador  $T_n$  inicial.

Tukey (1958), por sua vez, notou que as *pseudo-observações*  $(T_{n1}^*, \dots, T_{nn}^*)$  podem ser tratadas como independentes. Segundo ele, um estimador razoável para a variância de  $T_n$  é

$$\text{VAR}(T_n) = \frac{1}{n(n-1)} \sum_{i=1}^n (T_{ni}^* - T_n^*)^2.$$

Utilizando a expressão (3.16) para calcular  $\text{IF}(x_i, T, F_n)$  e no caso de  $\varepsilon = -\frac{1}{n-1}$  vem a relação

$$\text{IF}(x_i, T, F) \cong \frac{T\left(\left(1 - \frac{-1}{n-1}\right) F_n + \left(\frac{-1}{n-1}\right) \Delta_{x_i}\right) - T(F_n)}{\frac{-1}{n-1}} = T_{ni}^* - T_n,$$

demonstrando como o  $i$ -ésimo *pseudo-valor jackknifed*  $T_{ni}^*$  é um indicador da incidência do valor de  $x_i$  no resultado do estimador  $T_n$ , podendo assim ser considerado como uma versão discreta e empírica da *função de influência*.

### 3.5.5 Ponto de ruptura

Uma questão que parece ser pertinente colocar quando estamos a tratar de métodos robustos, que estão relacionados com a existência de “outliers”, é qual a percentagem de “gross errors” (ou “outliers”) que o procedimento estatístico tolera antes de dar resultados absurdos. Esta fracção  $\varepsilon^*$  é conhecida por “breakdown point”, termo que se traduziu por *ponto de ruptura*. Este conceito foi introduzido por Hodges (1967) e generalizado por Hampel (1968).

Existem várias definições de *ponto de ruptura*. Vamos apresentar uma versão assintótica e outra amostral.

Considerando a vizinhança de contaminação e o enviesamento assintótico máximo dado por  $b(\varepsilon) = \sup_{G \in GE(F, \varepsilon)} |T(G) - T(F)|$ , vem a seguinte definição:

**Definição 3.4 (breakdown point assintótico, Hampel, 1968)** O “breakdown point” assintótico de  $T$  é dado por

$$\varepsilon^* = \varepsilon^*(T, F) = \inf \{ \varepsilon : b(\varepsilon) = \infty \}. \quad (3.25)$$

Normalmente esta constante é independente da distribuição  $F$ .

Para a versão empírica considere-se um estimador  $T_n$  e uma amostra  $X = (x_1, \dots, x_n)$ . Agora tomem-se todas as possíveis amostras “corruptas”,  $X^*$ , obtidas substituindo as observações por valores arbitrários e considere-se  $\varepsilon = f/n$ . Representado o estimador com a amostra alterada por  $T_n^*$  o enviesamento máximo causado pela contaminação é então  $b_n(\varepsilon) = \sup |T_n - T_n^*|$ . Temos assim a seguinte definição:

**Definição 3.5 (breakdown point empírico, Donoho e Huber, 1983)** O “breakdown point” empírico do estimador  $T_n$  é dado por

$$\varepsilon_n^* = \varepsilon_n^*(T_n, F_n) = \min \left\{ \varepsilon = \frac{f}{n} : b_n(\varepsilon) = \infty \right\}. \quad (3.26)$$

A maior quantidade de “más” observações que um estimador razoável pode suportar é metade do número de observações redundantes, i.e. metade de  $(n - p)$  onde  $p$  é o número de parâmetros a estimar. Assim, resulta que

$$\varepsilon_{\max}^* = \frac{[(n - p)/2]}{n}, \quad (3.27)$$

onde  $[\cdot]$  representa a característica do número. À medida que  $n \rightarrow \infty$ ,  $\varepsilon_{\max}^*$  aproxima-se de 50%. Um estimador com “breakdown point” próximo de 50% é designado de estimador com *alto ponto de ruptura* (“high breakdown point estimator”).

Vejamos alguns exemplos.

## 1. Média

Considere-se o funcional média  $T_1(F) = \int x dF(x) = \mu(F)$ . Temos que

$$T_1(F_{x,\varepsilon}) = (1 - \varepsilon) \mu(F) + \varepsilon x,$$

portanto o modelo contaminado tem uma média que é uma soma ponderada da média da distribuição  $F$  com o ponto  $x$ . Uma vez que  $x$  pode variar livremente, para qualquer valor de  $\varepsilon$  positivo vem que  $T_1(F_{x,\varepsilon})$  é ilimitado em  $x$  e como consequência  $\varepsilon^* = 0$ .

Quanto à versão amostral temos que  $\varepsilon_n^* = 1/n$ , já que apenas uma observação é necessária para levar o estimador para fora de qualquer limite.

## 2. Mediana

Seja  $T_3(F) = F^{-1}\left(\frac{1}{2}\right)$  o funcional mediana. Para se encontrar o ponto de ruptura vamos fixar  $\varepsilon \leq 1/2$ . Então, para  $x$  suficientemente grande vem que  $F_{x,\varepsilon}^{-1}\left(\frac{1}{2}\right) = F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$  (ver função de influência da mediana pág. 57). Portanto,  $F_{x,\varepsilon}^{-1}\left(\frac{1}{2}\right)$  é limitado em  $x$ .

No entanto, para  $\varepsilon > \frac{1}{2}$  vem que  $F_{x,\varepsilon}^{-1}\left(\frac{1}{2}\right) > x$  sendo agora ilimitada em  $x$ . Resulta assim que  $\varepsilon^* = \frac{1}{2}$ , sendo um estimador com alto ponto de ruptura o que faz com que este seja um dos estimadores robustos de localização mais utilizados.

## 3.6 Estimadores de dispersão univariados

No modelo de escala,  $\{F_\theta, \theta \in \mathbb{R}^+\}$ , supõe-se que  $F_\theta(x) = F\left(\frac{x}{\theta}\right)$  e  $f_\theta(x) = \frac{1}{\theta}f\left(\frac{x}{\theta}\right)$ , onde  $\theta$  é o parâmetro de escala. Este modelo corresponde a uma restrição do modelo geral de localização e escala,  $F_{\mu,\theta}(x) = F\left(\frac{x-\mu}{\theta}\right)$ , com localização,  $\mu$ , supostamente conhecida e igual a zero.

Uma propriedade que os funcionais de escala devem verificar é a *equivariância em escala*, isto é,

$$\sigma\left(F\left(\frac{x}{a}\right)\right) = |a| \sigma(F(x)), \quad \text{para todo o } a \neq 0. \quad (3.28)$$

No caso de  $X \sim F$  é mais usual utilizar-se a notação  $\sigma(aX) = |a| \sigma(X)$  o que traduz a ideia que  $\sigma(X)$  representa  $\sigma(F)$ . Vamos assumir que os funcionais considerados são equivalentes aos estimadores correspondentes, pelo menos aproximadamente ou assintoticamente.

Na maioria das situações, mesmo que o parâmetro de interesse seja o de escala, o parâmetro de localização (perturbador) também é desconhecido. Assim, para se efectuar a estimação é necessário primeiro realizar uma translação de forma a centrar a variável. Considerando que  $\mu$  é o funcional de localização devemos então considerar  $X - \mu \sim F_0$  onde  $F_0$  é uma distribuição da mesma família paramétrica de  $F$ . Daí ser conveniente que o estimador de escala seja também *invariante em translação*, verificando-se a relação para uma amostra aleatória da variável aleatória  $X$ ,

$$s(aX_1 + b, \dots, aX_n + b) = |a| s(X_1, \dots, X_n), \quad (3.29)$$



onde  $a$  e  $b$  são constantes arbitrárias, com  $a \neq 0$  e  $s$  é um estimador de escala. Esta última relação, mais forte que a de (3.28), é usualmente conhecida por *equivariância afim* (“affine equivariance”).

Vamos agora ver os dois estimadores robustos de dispersão univariados que se utilizam nesta tese.

## 1. Desvio mediano absoluto (MAD)

Como vimos anteriormente, a mediana é um estimador de localização “bastante” robusto, tendo em conta a *sensibilidade a grandes erros* e o *ponto de ruptura*. Em contrapartida, no que diz respeito à escala, o estimador *desvio mediano absoluto em relação à mediana* (MAD) é o equivalente à mediana em termos das propriedades de robustez. Este é dado por

$$\text{MAD} = b \text{ med } |X_i - \text{med}(X_j)|, \quad (3.30)$$

onde  $\text{med}$  representa a mediana da amostra  $(X_1, \dots, X_n)$ . A constante  $b$  que intervém na expressão deste estimador é necessária para o tornar consistente para o parâmetro e modelo em causa.

Quando o parâmetro de dispersão  $\sigma$  é o desvio padrão de uma distribuição normal, o valor que se fixa para  $b$  é  $1.4836 \cong 1/\Phi^{-1}(3/4)$ . Quando  $F = \Phi$  a *função de influência* é

$$\text{IF}(x, \sigma_{\text{MAD}}, \Phi) = \begin{cases} \frac{b}{4\phi(\Phi^{-1}(\frac{1}{4}))}, & x < \Phi^{-1}(\frac{1}{4}) \vee x > \Phi^{-1}(\frac{3}{4}) \\ -\frac{b}{4\phi(\Phi^{-1}(\frac{1}{4}))}, & \Phi^{-1}(\frac{1}{4}) < x < \Phi^{-1}(\frac{3}{4}) \end{cases}, \quad (3.31)$$

onde  $\sigma_{\text{MAD}}$  representa o funcional relacionado com o estimador MAD.

Como se pode observar a *função de influência* é limitada. Considerando  $b = 1.4836$  vem  $\gamma^*(\text{MAD}, \Phi) \cong 1.167$  e  $\text{ASVAR}(\text{MAD}, \Phi) = (\gamma^*)^2 \cong 1.361$ , estamos assim na presença de um estimador *B-robusto* que goza ainda de *robustez qualitativa*. Além disso, o MAD é um estimador com *alto ponto de ruptura*,  $\varepsilon^* = \frac{1}{2}$ .

Este estimador tem-se tornado bastante popular devido à sua simplicidade e às boas propriedades de robustez que possui. O seu cálculo computacional utiliza um algoritmo que consome  $O(n)$  unidades de tempo. Daí que o MAD tenha sido um dos estimadores de escala univariados seleccionados para intervir no nosso estudo.

O MAD foi bastante promovido por Hampel (1974), embora este autor o atribua a Gauss (1880).

Este estimador é também muitas vezes utilizado como ponto inicial do processo iterativo de estimadores mais eficientes. O estudo de simulação de Andrews *et al.*

(1972) confirmou a vantagem de se utilizarem pontos iniciais robustos no processo computacional dos estimadores-M. O MAD é muitas vezes utilizado para este efeito por ser um estimador com *alto ponto de ruptura* e ancilar relativamente ao parâmetro de localização. Como realçou Huber (1981, pág 107), esta propriedade é de extrema importância, pois os estimadores-M de localização necessitam de um estimador de escala ancilar para os tornar *equivariantes*.

Em contrapartida, o MAD também tem algumas desvantagens. A sua eficiência no caso do modelo normal é bastante baixa, apenas de 37%. Além disso, o MAD só é indicado para distribuições simétricas. A sua construção assume uma visão simétrica da dispersão, uma vez que se procura o intervalo simétrico, em torno da mediana, que contém 50% das observações. Este procedimento não parece o mais indicado no caso da distribuição ser assimétrica.

## 2. Estimador-M de escala

Existem duas alternativas possíveis para se encontrar um estimador-M de escala quando a localização é desconhecida. Uma delas é escolher previamente um estimador para a localização,  $(T_n)$ , que é sempre fixo, e o estimador-M de escala  $(s_n)$  deverá satisfazer a equação implícita

$$\sum_{i=1}^n \chi \left( \frac{x_i - T_n}{s_n} \right) = 0.$$

Para obter uma solução numérica é necessário considerar estimativas iniciais para a localização (a qual se mantém fixa) e para a escala. Muitas vezes selecciona-se a mediana e o MAD como estimativas iniciais de localização e escala, respectivamente. A equação correspondente em termos de funcionais,  $T$  para a localização e  $\sigma$  para a escala, é  $\int \chi \left( \frac{x - T(F)}{\sigma(F)} \right) dF(x) = 0$ .

A outra hipótese é considerar um estimador-M simultâneo de localização-escala, onde ambas as estimativas iniciais (localização e escala) vão sendo actualizadas ao longo do processo de cálculo. Assim, qualquer par de estatísticas  $(T_n, s_n)$  que satisfaça as equações

$$\sum_{i=1}^n \psi \left( \frac{x_i - T_n}{s_n} \right) = 0, \quad (3.32)$$

$$\sum_{i=1}^n \chi \left( \frac{x_i - T_n}{s_n} \right) = 0, \quad (3.33)$$

é um estimador-M de localização-escala. Em termos de funcionais vem que

$$\int \psi \left( \frac{x - T(F)}{\sigma(F)} \right) dF(x) = 0, \quad (3.34)$$

$$\int \chi \left( \frac{x - T(F)}{\sigma(F)} \right) dF(x) = 0. \quad (3.35)$$

De forma análoga ao Exemplo 4 apresentado na Subsecção 5.2 deste capítulo (pág. 58), as *funções de influência* são encontradas substituindo  $F$  por  $F_{x,\varepsilon}$  em (3.34) e (3.35) e, posteriormente, avaliando as derivadas em ordem a  $\varepsilon$  no ponto  $\varepsilon = 0$ . Estas verificam o sistema seguinte

$$\text{IF}(x, T, F) \int \psi'(y) dF(x) + \text{IF}(x, \sigma, F) \int \psi'(y) y dF(x) = \psi(y) \sigma(F),$$

$$\text{IF}(x, T, F) \int \chi'(y) dF(x) + \text{IF}(x, \sigma, F) \int \chi'(y) y dF(x) = \chi(y) \sigma(F),$$

onde  $y = \left( \frac{x - T(F)}{\sigma(F)} \right)$ .

No caso de  $F$  ser simétrica,  $\psi$  uma função ímpar, e  $\chi$  uma função par, resulta como solução do sistema anterior,

$$\text{IF}(x, T, F) = \frac{\psi \left( \frac{x - T(F)}{\sigma(F)} \right) \sigma(F)}{\int \psi' \left( \frac{x - T(F)}{\sigma(F)} \right) dF(x)}, \quad (3.36)$$

$$\text{IF}(x, \sigma, F) = \frac{\chi \left( \frac{x - T(F)}{\sigma(F)} \right) \sigma(F)}{\int \chi' \left( \frac{x - T(F)}{\sigma(F)} \right) \left( \frac{x - T(F)}{\sigma(F)} \right) dF(x)}. \quad (3.37)$$

Contudo, esta variante do estimador-M tem a desvantagem de não manter o *alto ponto de ruptura* do estimador de escala inicial. Como alternativa, é muitas vezes recomendado o emprego de versões ponderadas a  $k$  passos. Com o estimador-M ponderado com um número fixo de  $k$  passos é possível combinar duas das características importantes de um estimador robusto, a alta eficiência assintótica com o *alto ponto de ruptura* (Rousseeuw e Croux, 1994).

Considerando  $(T_n^0, s_n^0)$  os estimadores iniciais de localização e de escala, os estimadores ponderados para a localização e para a escala, podem ser definidos por

$$T_n^{(k+1)} = \frac{\sum_{i=1}^n w_1(u_i^{(k)}) x_i}{\sum_{i=1}^n w_1(u_i^{(k)})}, \quad (3.38)$$

$$(s_n^{(k+1)})^2 = \frac{\sum_{i=1}^n w_2(u_i^{(k)}) (x_i - T_n^{(k)})^2}{n\beta}, \quad (3.39)$$

com  $u_i^{(k)} = \frac{x_i - T_n^{(k)}}{s_n^{(k)}}$ , onde  $w_j(u_i)$  para  $j = 1, 2$  são pesos convenientes e  $\beta$  é uma constante que torna o estimador consistente para o parâmetro de interesse, sob a distribuição normal.

O estimador-M seleccionado para o nosso trabalho foi o mesmo de Boente e Orellana (2001), dado que um dos nossos objectivos é confrontar os resultados obtidos nesta tese com os alcançados pelas referidas autoras.

A escolha de Boente e Orellana (2001) para estimador inicial de dispersão (que se actualiza com pesos ao longo do processo iterativo) foi o estimador MAD com *alto ponto de ruptura*, e assumindo como modelo central o modelo normal fixou-se a constante  $b = 1.4826$  (ver (3.30)). Para manter o *alto ponto de ruptura* usou-se a versão ponderada, com 10 passos, sendo as funções  $\psi$  e  $\chi$ , para a localização e escala, respectivamente

$$\psi(u) = \min\left(\frac{u}{1.345}, 1\right), \quad (3.40)$$

$$\chi(u) = \min\left(\frac{u^2}{c^2}, 1\right) - \beta = \min\left(\frac{u^2}{(1.041)^2}, 1\right) - \frac{1}{2}, \quad (3.41)$$

o que conduz a um estimador com eficiência de 0.509, sob o modelo normal.

### 3.7 Algumas generalidades sobre estimação multivariada

Nesta secção vamos apresentar alguns tópicos da estimação robusta multivariada, essenciais para o nosso trabalho, uma vez que o modelo em estudo, modelo das componentes principais comuns, é da área da análise multivariada. Como a metodologia que se desenvolveu deu especial relevância à abordagem infinitesimal de Hampel, em particular à *função de influência*, vamos apresentar alguns dos conceitos introduzidos por Hampel, mas para o caso multivariado.

Antes de mais, convém realçar que se desejam algumas propriedades para os estimadores robustos multivariados de localização e de dispersão. Entre elas podemos citar a *equivariância para transformações afins*.

**Definição 3.6 (estimadores equivariantes afins)** Seja  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  uma amostra aleatória de um vector aleatório de  $\mathbb{R}^p$  e  $\mathbf{t}_n = \mathbf{t}(\mathbf{X}_1, \dots, \mathbf{X}_n) \in \mathbb{R}^p$  um estimador do vector de localização e  $\mathbf{V}_n = \mathbf{V}(\mathbf{X}_1, \dots, \mathbf{X}_n) \in PDS(p)$  um estimador da matriz de dispersão. Dizemos que  $\mathbf{t}_n$  e  $\mathbf{V}_n$  são *equivariantes para transformações afins* se

para qualquer matriz  $\mathbf{A}_{(p \times p)}$  não singular e qualquer vector  $\mathbf{b} \in \mathbb{R}^p$ , se tem

$$\mathbf{t}(\mathbf{A}\mathbf{X}_1 + \mathbf{b}, \dots, \mathbf{A}\mathbf{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{t}(\mathbf{X}_1, \dots, \mathbf{X}_n) + \mathbf{b}, \quad (3.42)$$

$$\mathbf{V}(\mathbf{A}\mathbf{X}_1 + \mathbf{b}, \dots, \mathbf{A}\mathbf{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{V}(\mathbf{X}_1, \dots, \mathbf{X}_n)\mathbf{A}^T. \quad (3.43)$$

Note-se que (3.43) não é mais que a generalização da condição (3.29).

Estas condições são bastante naturais e são muitos os estimadores multivariados de localização e dispersão que as verificam, robustos ou clássicos (não robustos). Neste último caso temos  $\bar{\mathbf{X}}$  e  $\mathbf{S}$ , o vector média amostral e a matriz de covariâncias amostral, respectivamente, que são os dois estimadores clássicos mais utilizados.

Por vezes, esta definição é relaxada considerando  $\mathbf{A}$  uma matriz ortogonal. Neste caso, diz-se que  $\mathbf{t}$  e  $\mathbf{V}$  possuem *equivariância ortogonal*.

Para os funcionais, deseja-se que estes tenham também esta característica. Considere-se que  $\mathbf{t} = \mathbf{t}(F)$  é o funcional vector de localização e  $\mathbf{V} = \mathbf{V}(F)$  o funcional matriz de dispersão ambos na classe de distribuições  $\mathcal{F}$  a qual contém  $F$ , a “verdadeira” distribuição de probabilidade, por exemplo a família de distribuições paramétricas  $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$  com  $\Theta \subset \mathbb{R}^m$ .

Dizemos que os funcionais  $\mathbf{t}$  e  $\mathbf{V}$  são *equivariantes para transformações afins* se para qualquer matriz  $\mathbf{A}_{(p \times p)}$  não singular e qualquer vector  $\mathbf{b} \in \mathbb{R}^p$ , se tem

$$\mathbf{t}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbf{t}(\mathbf{X}) + \mathbf{b}, \quad (3.44)$$

$$\mathbf{V}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbf{V}(\mathbf{X})\mathbf{A}^T, \quad (3.45)$$

onde  $\mathbf{X}$  representa um vector aleatório de  $\mathbb{R}^p$ .

Para além destas propriedades é desejável que os funcionais multivariados e os respectivos estimadores equivalentes tenham boas propriedades de robustez. Vejamos então as seguintes propriedades.

Considere-se a distribuição contaminada  $F_{\mathbf{x},\varepsilon} = (1 - \varepsilon)F + \varepsilon\Delta_{\mathbf{x}}$  onde  $\Delta_{\mathbf{x}}$  é a distribuição de Dirac que coloca toda a sua massa em  $\mathbf{x} \in \mathbb{R}^p$ . Temos a seguinte definição.

**Definição 3.7 (função de influência  $p$ -dimensional)** A *função de influência* (IF) do funcional  $\mathbf{T}$  em  $F$  é dada por

$$\text{IF}(\mathbf{x}, \mathbf{T}, F) = \lim_{\varepsilon \rightarrow 0} \left[ \frac{\mathbf{T}(F_{\mathbf{x},\varepsilon}) - \mathbf{T}(F)}{\varepsilon} \right]. \quad (3.46)$$

com  $\mathbf{x} \in \Omega$  e para o qual este limite existe.

Sob certas condições de regularidade o valor esperado da *função de influência* é um vector nulo, isto é

$$\int \text{IF}(\mathbf{x}, \mathbf{T}, F) dF(\mathbf{x}) = \mathbf{0}_m, \quad (3.47)$$

e pelo teorema do limite central vem ainda que  $[n^{1/2}(\mathbf{T}(F_n) - \mathbf{T}(F))]$  se distribui assintoticamente segundo uma normal multivariada com média  $\mathbf{0}_m$  e matriz de covariâncias assintótica (que designaremos por variância assintótica, ASVAR) igual a

$$\text{ASVAR}(\mathbf{T}, F) = E \left[ \text{IF}(\mathbf{X}, \mathbf{T}, F) \text{IF}(\mathbf{X}, \mathbf{T}, F)^T \right]. \quad (3.48)$$

Ainda se verifica a seguinte relação (generalização de 3.22) para a variância assintótica,

$$\text{ASVAR}(\mathbf{T}, F) \geq \mathbf{I}(\boldsymbol{\theta})^{-1}, \quad (3.49)$$

onde “ $\geq$ ” representa a relação de ordem (forte) entre matrizes e  $\mathbf{I}(\boldsymbol{\theta})$  a matriz de informação de Fisher.

Tal como no caso unidimensional a *sensibilidade a grandes erros* também se pode definir, mas substituindo o módulo pela norma euclidiana. Esta constante por vezes é preterida em favor de variantes standardizadas, quer por  $\text{ASVAR}(\mathbf{T}, F)$  (caso exista), quer por  $\mathbf{I}(\boldsymbol{\theta})^{-1}$  (caso exista). Com isto resolve-se o problema da falta de equivariância para transformações de escala presente na *sensibilidade a grandes erros*.

A alteração sofrida na definição do *ponto de ruptura* é a mesma da *sensibilidade a grandes erros*, isto é, o módulo é substituído pela norma euclidiana.

Convém realçar que no caso do funcional ser uma matriz  $\mathbf{V} \in PDS(p)$ , como acontece com o funcional matriz de dispersão, é possível usar outra notação para apresentar a *função de influência*. A IF tem neste caso  $p(p+1)/2$  componentes relevantes, que é o número de elementos da matriz triangular (superior ou inferior) relacionada com  $\mathbf{V}$ . Uma possibilidade para apresentar os elementos desta matriz triangular é aplicar o operador  $\text{vech}$  (ver Apêndice A) à matriz  $\mathbf{V}$ .

### 3.8 Estimadores de dispersão multivariados

Nesta subsecção vamos fazer referência aos estimadores robustos de dispersão multivariados que estão relacionados com este trabalho por:

1. serem utilizados no nosso trabalho como estimadores da matriz de covariâncias,
2. serem utilizados nos trabalhos com os quais o nosso será comparado.

Do grupo **1.** considerou-se o estimador que passamos a descrever.

### 1. Matriz de Covariâncias de Determinante Mínimo (MCD)

O estimador “Minimum Covariance Determinant”, conhecido na literatura pela abreviatura MCD, foi proposto por Rousseeuw (1984). A ideia para o construir é bastante simples. Como o nome sugere, procura-se encontrar o subconjunto dos dados (de cardinalidade  $h$ )  $\mathcal{S}_h = \{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_h}\}$ , com  $p + 1 \leq h \leq n$ , cuja matriz de covariâncias tem o menor determinante. A estimativa MCD de localização é então definida como

$$\mathbf{t}_n = \frac{1}{h} \sum_{j=1}^h \mathbf{x}_{i_j}, \quad (3.50)$$

e a da matriz de dispersão por

$$\mathbf{V}_n = c_p \frac{1}{h} \sum_{j=1}^h (\mathbf{x}_{i_j} - \mathbf{t}_n) (\mathbf{x}_{i_j} - \mathbf{t}_n)^T, \quad (3.51)$$

onde  $c_p$  é um factor de consistência,  $\mathbf{x}_{i_j}$  e  $\mathbf{t}_n$  são vectores de  $\mathbb{R}^p$ .

É comum considerar-se  $h = [(n + p + 1)/2]$  porque com este valor obtém-se um estimador com *alto ponto de ruptura* (Lopuhaä e Rousseeuw, 1991). No caso multivariado, o *ponto de ruptura* pode ser relacionado com os valores próprios da matriz de covariâncias. A “ruína” da matriz de covariâncias acontece quando o maior valor próprio tende para infinito (“explosão”) ou o menor valor próprio tende para zero (“implosão”). O *ponto de ruptura* é o ponto imediatamente anterior à “ruína”.

Outra alternativa para  $h$ , também muito utilizada, é o valor  $[0.75n]$ . Com este  $h$  consegue-se obter um bom compromisso entre eficiência/estabilidade e *ponto de ruptura*.

Este estimador é *equivariante para transformações afins* e é, como foi constatado por Croux e Haesbroeck (1999a), particularmente indicado para estimar o parâmetro  $\Sigma$  de uma distribuição elipticamente simétrica. Neste caso, a f.d.p.  $f$  é da forma

$$f_{\boldsymbol{\mu}, \Sigma}(\mathbf{x}) = |\Sigma|^{-1/2} f\left(\Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (3.52)$$

com matriz de dispersão  $\Sigma = \mathbf{C}\mathbf{C}^T \in PDS(p)$  e  $\boldsymbol{\mu}$  o vector localização e  $f$  é qualquer densidade de  $\mathbb{R}^p$ . Para este tipo de distribuição tem-se sempre (dada a

simetria)  $\boldsymbol{\mu} = E[\mathbf{X}]$  mas  $\boldsymbol{\Sigma}$  não coincide, em geral com  $\text{VAR}[\mathbf{X}]$ , garantindo-se apenas que  $\boldsymbol{\Sigma} = a\text{VAR}[\mathbf{X}]$ , com  $a \in \mathbb{R}^+$ . Por esta razão chama-se a  $\boldsymbol{\Sigma}$  matriz de pseudo-covariâncias ou matriz de dispersão.<sup>1</sup> Em face desta proporcionalidade as estruturas de covariâncias descritas no Capítulo 2 aplicam-se igualmente às matrizes de dispersão. Daqui em diante passaremos a usar, sempre que possível, esta designação.

Tal como acontece com o estimador-M de escala, a eficiência do MCD também aumenta com a versão ponderada. Nesta versão, o estimador é habitualmente referido por RMCD (Rousseeuw, 1985).

Considerando  $(\mathbf{t}_n^0, \mathbf{V}_n^0)$  os estimadores de localização e de dispersão iniciais, os estimadores ponderados a um passo para a localização e para a dispersão, são definidos com

$$\mathbf{t}_n^1 = \frac{\sum_{i=1}^n w_i \mathbf{X}_i}{\sum_{i=1}^n w_i}, \quad (3.53)$$

$$\mathbf{V}_n^1 = c_1 \frac{\sum_{i=1}^n w_i (\mathbf{X}_i - \mathbf{t}_n^1) (\mathbf{X}_i - \mathbf{t}_n^1)^T}{\sum_{i=1}^n w_i}. \quad (3.54)$$

Os pesos atribuídos dependem dos estimadores iniciais,

$$w_i = w \left( (\mathbf{X}_i - \mathbf{t}_n^0)^T (\mathbf{V}_n^0)^{-1} (\mathbf{X}_i - \mathbf{t}_n^0) \right),$$

sendo  $w$  uma aplicação de  $\mathbb{R}^+$  em  $\mathbb{R}$ .

Habitualmente os pesos considerados no passo posterior são,

$$w(y) = I_{[0, q_\delta]}(y) \quad \text{com } q_\delta = G^{-1}(1 - \delta), \quad (3.55)$$

e  $G(u) = P(\mathbf{X}^T \mathbf{X} \leq u)$ .

O funcional MCD de escala ponderado a um passo,  $\mathbf{V}^1 = \mathbf{V}(F)^1$ , com os pesos indicados em (3.55), tem *função de influência*

$$\begin{aligned} \text{IF}(\mathbf{x}, \mathbf{V}^1, F) &= \frac{(1-\delta)c_1^{-1} + 2d_3}{(1-\delta)c_1^{-1}} \left( \text{IF}(\mathbf{x}, \mathbf{V}^0, F) + \frac{1}{2} \text{tr}(\text{IF}(\mathbf{x}, \mathbf{V}^0, F)) \mathbf{I} \right) + \\ &+ \frac{1}{(1-\delta)c_1^{-1}} \mathbf{I}(\mathbf{x} \mathbf{x}^T \leq q_\delta) \mathbf{x} \mathbf{x}^T - \mathbf{I}, \end{aligned} \quad (3.56)$$

onde  $d_3$  é uma constante e  $\text{tr}$  representa o traço da matriz. Este resultado encontra-se no trabalho de Croux e Haesbroeck (1999a) onde também são apresentados gráficos desta função.

---

<sup>1</sup>Note-se que para a distribuição normal se tem  $a = 1$ .



O MCD tem ainda a vantagem de ter a taxa de convergência normal, i. e.  $O(n^{\frac{1}{2}})$ , o que não acontece por exemplo com o estimador “minimum volume ellipsoid” (MVE), como verificou Davies (1992). A distribuição do MCD de localização é assintoticamente normal, o que é também uma propriedade desejável.

Com o algoritmo proposto por Rousseeuw e van Driessen (1999), designado por FAST-MCD, o seu cálculo computacional é extremamente rápido mesmo em problemas de grandes dimensões.

Foram as razões acima apresentadas que nos levaram a optar por este estimador. Utilizamos a versão RMCD, com  $h = [0.75n]$  atingindo-se um estimador com um *ponto de ruptura* de 25%.

Do grupo 2. de estimadores, aqueles cuja escolha nos transcende, fazem parte o estimador-M da matriz de dispersão e o estimador de Stahel–Donoho, que passamos a descrever de seguida.

## 2. Estimador-M multivariado de localização–dispersão

Os estimadores-M multivariados de localização–dispersão, *equivariantes para transformações afins*, foram propostos por Maronna (1976), para distribuições elipticamente simétricas, como uma generalização dos estimadores de máxima verosimilhança. Considerando  $\mathbf{X}_i$ ,  $i = 1, \dots, n$ , vectores aleatórios de  $\mathbb{R}^p$ , com vector média  $\boldsymbol{\mu} \in \mathbb{R}^p$  e matriz de dispersão  $\boldsymbol{\Sigma} \in PDS(p)$ , os estimadores-M  $(\mathbf{t}_n, \mathbf{V}_n)$  são as soluções do sistema de equações implícitas

$$n^{-1} \sum_{i=1}^n u_1 [d(\mathbf{X}_i, \mathbf{t}_n, \mathbf{V}_n)] (\mathbf{X}_i - \mathbf{t}_n) = \mathbf{0}_p, \quad (3.57)$$

$$n^{-1} \sum_{i=1}^n u_2 [d^2(\mathbf{X}_i, \mathbf{t}_n, \mathbf{V}_n)] (\mathbf{X}_i - \mathbf{t}_n) (\mathbf{X}_i - \mathbf{t}_n)^T = \mathbf{V}_n, \quad (3.58)$$

onde  $u_i$  são funções genéricas e  $d^2(\mathbf{X}, \mathbf{t}_n, \mathbf{V}_n) = (\mathbf{X} - \mathbf{t}_n)^T \mathbf{V}_n^{-1} (\mathbf{X} - \mathbf{t}_n)$ .

Como facilmente se observa, com  $u_1 = u_2 = 1$  resultam os EMV da normal multivariada. Considerando versões ponderadas da média amostral e da matriz de covariâncias amostral, onde os pesos são determinados pelos dados, usando uma versão adaptada da distância de Mahalanobis, temos uma interpretação bastante intuitiva para os estimadores-M. Esta ideia pode ser utilizada na construção do processo iterativo. A taxa de convergência deste processo foi estudada por Kent e Tyler (1991).

A existência e unicidade das soluções de (3.57) e (3.58) sob certas condições, assim como a consistência e normalidade assintótica, foram obtidas por Maronna (1976).

Os funcionais  $\mathbf{t}(F)$  e  $\mathbf{V}(F)$  equivalentes aos estimadores  $\mathbf{t}_n$  e  $\mathbf{V}_n$  são definidos de forma análoga e consistente como a solução de,

$$E_F \{u_1 [d(\mathbf{X}, \mathbf{t}(F), \mathbf{V}(F))] (\mathbf{X} - \mathbf{t}(F))\} = \mathbf{0}_p, \quad (3.59)$$

$$E_F \left\{ u_1 \left[ d^2(\mathbf{X}, \mathbf{t}(F), \mathbf{V}(F)) \right] (\mathbf{X} - \mathbf{t}(F)) (\mathbf{X} - \mathbf{t}(F))^T \right\} = \mathbf{V}(F). \quad (3.60)$$

Uma característica pouco atractiva dos estimadores-M é o seu baixo *ponto de ruptura* em problemas de grande dimensionalidade.

Maronna (1976) calculou o limite superior para o *ponto de ruptura*, sendo este valor  $\varepsilon^* \leq \min(1/K_2, 1 - p/K_2)$  com  $K_2 = \sup_d \{d^2 u_2(d^2)\}$ . Como  $K_2 > p$  resulta que, para qualquer estimador-M,  $\varepsilon^* \leq (p+1)^{-1}$ . Tyler (1986) mostrou que o *ponto de ruptura* exacto do estimador-M multivariado é de facto o  $\min(1/K_2, 1 - p/K_2)$ . Assim, o valor *ponto de ruptura* vai diminuindo à medida que  $p$  aumenta.

As estimativas-M obtêm-se com o seguinte algoritmo. Considerando que  $(\mathbf{t}^0, \mathbf{V}^0)$  são as estimativas iniciais, então as estimativas de ordem  $(k+1)$  calculam-se com

$$\mathbf{t}_n^{(k+1)} = \frac{\sum_{i=1}^n u_1 \left( d_i^{(k)} \right) \mathbf{x}_i}{\sum_{i=1}^n u_1 \left( d_i^{(k)} \right)},$$

e

$$\mathbf{V}_n^{(k+1)} = \frac{\sum_{i=1}^n u_2 \left( (d_i^2)^{(k)} \right) (\mathbf{x}_i - \mathbf{t}_n) (\mathbf{x}_i - \mathbf{t}_n)^T}{n},$$

com  $d_i^{(k)} = d(\mathbf{X}_i, \mathbf{t}_n^{(k)}, \mathbf{V}_n^{(k)})$ .

Para se escrever a *função de influência* do funcional de dispersão, convém rescrever a equação (3.60) com o operador vech, pois é importante excluir os  $p(p-1)/2$  elementos em duplicado. Temos assim,

$$E_F [\psi(\mathbf{x}, \text{vech}(\mathbf{V}))] = \mathbf{0}_p,$$

com

$$\psi(\mathbf{x}, \text{vech}(\mathbf{V})) = u_2(\mathbf{x}^T \mathbf{V}^{-1} \mathbf{x}) \text{vech}(\mathbf{x} \mathbf{x}^T) - \text{vech}(\mathbf{V}),$$

no caso de  $\boldsymbol{\mu} = \mathbf{0}_p$ , para simplificar a notação. Então a *função de influência* é

$$\text{IF}(\mathbf{x}, \text{vech}(\mathbf{V}), F) = \mathbf{M}^{-1} \psi(\mathbf{x}, \text{vech}(\mathbf{V})), \quad (3.61)$$

com

$$\mathbf{M}^{-1} = \int \left[ \frac{\partial \psi(\mathbf{x}, \text{vech}(\mathbf{V}))}{\partial \text{vech}(\mathbf{V})} \right] dF(\mathbf{x}).$$

Embora esta função tenha um aspecto complicado, é fácil verificar que é limitada desde que  $\psi$  seja limitada.

O estimador-M multivariado que citaremos neste trabalho, consta nos trabalhos de Orellana (1999) e de Boente e Orellana (2001). As autoras tomaram para pesos a proposta 2 multivariada de Huber, isto é,  $u_i(s) = \psi_i(s)/s$ ,  $i = 1, 2$ . Consideraram  $\psi_1(s) = \psi(s, k_1)$  e  $\psi_2(s) = \psi(s, k_2^2)/\beta$ , com  $\psi(z, k) = \max(-k, \min(z, k))$ ,  $k_1$  e  $k_2$  constantes positivas e  $\beta = E\psi(\|X\|^2, k_2)$ . Atribuíram a  $k_1$  e  $k_2$  o valor 2.7649, o que implica  $\beta$  igual a 0.9363 e *ponto de ruptura* de aproximadamente 12%. Estes valores foram obtidos por Boente (1983), para se alcançar uma eficiência assintótica de 95% para o modelo normal.

### 3. Estimador de Stahel–Donoho (SDE)

Na procura de um estimador de localização e dispersão que combinasse ambas as propriedades, *equivariância para transformações afins* e *alto ponto de ruptura*, Stahel (1981) e Donoho (1982) propuseram independentemente o primeiro estimador robusto com estas duas propriedades. A ideia para a construção do SDE está relacionada com a identificação de “outliers”. Estes autores consideraram que um “outlier” multivariado deveria aparecer como “outlier” univariado nalguma projecção dos dados. Isto conduz à definição de uma medida de *discordância* (“outlyingness”) denotada por  $r_i$  para cada observação  $\mathbf{x}_i$ , nas projecções consideradas. O SDE de localização e dispersão é então definido como a média ponderada e a matriz de covariâncias ponderada, onde a cada observação é atribuído um peso função de  $r_i$ .

Considere-se  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$  uma amostra de  $n$  vectores aleatórios de  $\mathbb{R}^p$ , e sejam  $t$  e  $s$  estimadores univariados *equivariantes* de localização e escala, respectivamente. Para cada  $\mathbf{y} \in \mathbb{R}^p$  a *discordância* é,

$$r(\mathbf{y}, \mathbf{X}) = \sup_{\mathbf{a}} r_1(\mathbf{y}, \mathbf{a}, \mathbf{X}) , \quad (3.62)$$

onde,

$$r_1(\mathbf{y}, \mathbf{a}, \mathbf{X}) = |\mathbf{a}^T \mathbf{y} - t(\mathbf{a}^T \mathbf{X})| / s(\mathbf{a}^T \mathbf{X}) , \quad (3.63)$$

com  $\mathbf{a} \in \mathcal{S}_p = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1\}$ .

Dada uma função  $w : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  o SDE de localização e dispersão  $(\mathbf{t}_n, \mathbf{V}_n)$  é definido por

$$\mathbf{t}_n = \frac{\sum_{i=1}^n w_i \mathbf{X}_i}{\sum_{i=1}^n w_i} , \quad (3.64)$$

$$\mathbf{V}_n = \frac{\sum_{i=1}^n w_i (\mathbf{X}_i - \mathbf{t}_n) (\mathbf{X}_i - \mathbf{t}_n)^T}{\sum_{i=1}^n w_i}, \quad (3.65)$$

com  $w_i = w(r(\mathbf{X}_i, \mathbf{X}))$ .

Dado os pesos serem definidos com projecções, este estimador é designado por estimador de projecção.

Stahel (1981) mostrou que quando  $t$  e  $s$  são estimadores com *alto ponto de ruptura* assintótico o estimador SDE também tem esta característica.

Tyler (1994) indicou quais as condições em que o *ponto de ruptura* empírico do SDE é  $[(n - p + 1)/2]/n$ , valor máximo do *ponto de ruptura* empírico para qualquer estatística de dispersão *equivariante afim*, com  $\mathbf{X}$  em *posição geral*<sup>2</sup> e  $n \geq p + 1$ .

Zuo *et al.* (2001) provaram que no caso da função dos pesos,  $w$ , ser contínua e diferenciável a distribuição assintótica do estimador de localização é normal. Para este caso particular, Zuo *et al.* (2001) também indicaram a forma da *função de influência* do funcional de localização equivalente ao SDE. Para isso, consideraram  $w$  uma função peso (contínua) e o funcional  $p$ -variado

$$\mathbf{t}(F) = \frac{\int w(PD(\mathbf{x}, F)) \mathbf{x} dF(\mathbf{x})}{\int w(PD(\mathbf{x}, F)) dF(\mathbf{x})}, \quad (3.66)$$

o qual é Fisher-consistente para um parâmetro de localização  $\boldsymbol{\mu}$ . Na expressão anterior  $PD$  refere uma profundidade em projecção (“projection depth”).

A “projection depth” de um ponto  $\mathbf{x} \in \mathbb{R}^p$  com respeito a uma distribuição  $F$  é definida por

$$PD(\mathbf{x}, F) = \frac{1}{1 + O(\mathbf{x}, F)},$$

onde

$$O(\mathbf{x}, F) = \sup_{\mathbf{a} \in \mathcal{S}_p} \frac{|\mathbf{a}^T \mathbf{x} - \mu(F[\mathbf{a}])|}{\sigma(F[\mathbf{a}])}, \quad (3.67)$$

sendo  $F[\mathbf{a}]$  a distribuição de  $\mathbf{a}^T \mathbf{X}$ . Quando  $F$  é a função de distribuição empírica obtém-se (3.62).

Então, estes autores provaram que

$$\sqrt{n}(\mathbf{t}(F_n) - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, VAR(K\boldsymbol{\mu}(\mathbf{X}))),$$

---

<sup>2</sup>Neste caso não existem mais de  $p$  pontos em qualquer subespaço *afim* de dimensionalidade  $(p - 1)$ .

com

$$K_{\boldsymbol{\mu}}(\mathbf{x}) = \frac{\int (\mathbf{y} - \boldsymbol{\mu}) w'(PD(\mathbf{y}, F)) f(\mathbf{y}, \mathbf{x}) dF(\mathbf{y}) + (\mathbf{x} - \boldsymbol{\mu}) w(PD(\mathbf{y}, F))}{\int w(PD(\mathbf{y}, F)) dF(\mathbf{x})},$$

e

$$f(\mathbf{y}, \mathbf{x}) = \frac{O(\mathbf{y}, F) f_2(\mathbf{x}, \mathbf{u}(\mathbf{y})) + f_1(\mathbf{x}, \mathbf{u}(\mathbf{y}))}{\sigma(F[\mathbf{u}(\mathbf{y})]) (1 + O(\mathbf{y}, F))^2},$$

onde  $w'$  é a função derivada de  $w$  que também deve ser contínua,  $f_1$  e  $f_2$  as *funções de influência* dos funcionais univariados  $\mu(F)$  e  $\sigma(F)$  e  $\mathbf{u}(\mathbf{x})$  a direcção em que (3.67) atinge o supremo.

A partir da forma de  $K_{\boldsymbol{\mu}}(\mathbf{x})$  Zuo *et al.* (2001) concluíram que caso os funcionais  $\mu(\cdot)$  e  $\sigma(\cdot)$  tenham *funções de influência* limitadas então a *função de influência* do funcional do tipo **t** é limitada. Como o funcional relacionado com o SDE é do tipo (3.66), considerando pesos contínuos e diferenciáveis, resulta uma *função de influência* limitada. Assim, nesse caso particular o estimador de localização Stahel–Donoho é *B-robusto*.

Recentemente, Gervini (2002), apresentou a expressão explícita da *função de influência* do funcional equivalente ao SDE da matriz de dispersão, admitindo determinadas condições de regularidade e para o caso particular de distribuições elipticamente simétricas. Representando por  $\mathbf{V} = \mathbf{V}(F)$  o funcional Stahel–Donoho de dispersão, a sua *função de influência* é

$$\begin{aligned} \text{IF}(\mathbf{x}, \mathbf{V}, F) &= \left\{ \frac{c_2(F)}{c_0(F)} g_2(\|\mathbf{x}\|) + \frac{w(\|\mathbf{x}\|^2/s_0^2) \|\mathbf{x}\|^2}{c_0(F)} \right\} \left\{ \frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^2} - \frac{\mathbf{I}}{p} \right\} + \\ &+ \left\{ \frac{c_2(F)}{c_0(F)} g_3(\|\mathbf{x}\|) + \frac{w(\|\mathbf{x}\|^2/s_0^2) \|\mathbf{x}\|^2 - c_3(F)}{c_0(F)} \right\} \frac{\mathbf{I}}{p}, \end{aligned} \quad (3.68)$$

onde  $g_2(t)$  e  $g_3(t)$  são funções reais e  $c_0(F)$ ,  $c_2(F)$  e  $c_3(F)$  referem constantes. Note-se que uma escolha adequada da função  $w$  pode tornar este estimador *B-robusto*.

A ordem de convergência deste estimador é a desejável,  $O(n^{\frac{1}{2}})$ .

Contudo, este estimador apresenta alguns problemas relacionados com o seu cálculo. Para dimensão 2 é possível usar um grelha suficientemente fina de  $\mathcal{S}_2$ , mas em dimensões superiores não existe um método fácil para se obter  $r$ .

Stahel (1981) propôs um algoritmo baseado em subamostras (de dimensão  $m$ ) para calcular uma versão aproximada de  $r$ . Maronna e Yohai (1995) estudaram diferentes alternativas para a selecção da dimensão e o número de subamostras a considerar, de forma a garantir que este estimador aproximado tenha um *ponto de ruptura* próximo

ao do SDE original. Mas os resultados só foram satisfatórios para amostras pequenas em dimensões baixas, no máximo  $n = 30$  e  $p = 6$ .

O estimador Stahel–Donoho que vamos referir e utilizar no nosso trabalho, foi o considerado por Orellana (1999) e Boente e Orellana (2001). Este foi construído considerando para estimadores univariados de localização e escala a mediana e o MAD (normalizado com a constante  $b = 1.4836$ ), respectivamente.

A função de pesos considerada foi  $w(r) = I(r)_{]0,c]} + (c/r)^q I(r)_{]c,+\infty[}$  onde  $I[.]$  é a função indicatriz. Atribui-se  $q = 2$  e  $c = \sqrt{\chi_p^2(0.95)}$ , de modo a obter um estimador eficiente tanto para amostras normais multivariadas como para distribuições com caudas pesadas (Cauchy multivariada, Maronna e Yohai, 1995). Note-se no entanto que com esta função de pesos não são aplicáveis os resultados de Zuo *et al.* (2001).

A medida de *discordância*  $r$  foi calculada com subamostras com  $m = 1000$ .

As ideias apresentadas neste capítulo servirão de base para o desenvolvimento desta tese, dando-se especial relevo à abordagem infinitesimal de Hampel.



## Capítulo 4

# Estimação Robusta das Componentes Principais Comuns

### 4.1 Introdução

A estimação robusta em *componentes principais comuns* é um tema bastante recente. Em “Common principal components and related models” Flury (1988, págs. 161–162) sugere esta área para investigação futura. No entanto, um ano antes, Keramidas *et al.* (1987) quando comparavam graficamente as *componentes principais* de vários grupos, também fizeram um breve apontamento a este tópico. Estes autores, como vimos no capítulo anterior, preocuparam-se em comparar as *componentes principais* de vários grupos, sugerindo para esse efeito o uso de um vector “típico”. Este vector permite avaliar os ângulos entre os vectores próprios de vários grupos. Computacionalmente o vector “típico” é calculado como  $\max_{\bar{\mathbf{x}}_{ij}} \bar{\mathbf{x}}_{ij}^T \mathbf{E} \bar{\mathbf{x}}_{ij}$ ,  $i =$

$1, \dots, k$ ,  $j = 1, \dots, p$ , onde  $\mathbf{E} = \sum_{i=1}^k \mathbf{x}_{ij} \mathbf{x}_{ij}^T$  e  $\bar{\mathbf{x}}_{ij}$  são vectores unitários. Portanto, o vector “típico”  $\bar{\mathbf{x}}_{ij}$  não é mais que o vector próprio associado ao maior valor próprio de  $\mathbf{E}$ . Estes autores notaram que este cálculo seria afectado pela presença de observações amostrais discordantes. Nestas situações, os vectores próprios amostrais dariam resultados anómalos. Assim, para tentar evitar esta situação Keramidas *et al.* (1987) sugeriam estimar o vector “típico” com o estimador “ $\alpha$ -trimmed mean”, Tukey (1960). Com este estimador robusto as observações mais discordantes do padrão geral não influenciam tanto o processo de estimação.

Schott, no trabalho de 1991 intitulado “Some tests for common principal components subspaces in several groups”, também afluou o problema da estimação robusta nas



CPC. Como vimos no Capítulo 2, Subsecção 3.4, este autor preocupou-se em construir um novo teste de hipóteses (aproximado) para avaliar a validade do subespaço gerado por  $q \leq p$  *componentes principais* ser comum aos  $k$  grupos. Mas, como o autor refere, quando nos deparamos com observações “outliers” na amostra, pode ser conveniente calcular os vectores próprios de uma matriz de dispersão robusta. Schott (1991) sugere então, substituir a matriz de covariâncias clássica por uma matriz de dispersão robusta estimada através do estimador-M multivariado proposto por Tyler (1987).

Todavia, foi só no ano de 1992 que Novi Inverardi e Flury desenvolveram um trabalho completo no tópico da estimação robusta em *componentes principais comuns*. Este tema tem assim cerca de 10 anos de existência. Porventura esta será uma das razões que justifica o reduzido número de publicações existentes.

A actualidade deste tópico, a sua relevância na área da estatística, a escassez de publicações, foram as principais razões que nos levaram a desenvolver esta investigação. Neste capítulo iremos apresentar alguns dos resultados que obtivemos. Para isso, organizámos o capítulo do seguinte modo.

Na Secção 2 apresentamos um resumo dos três trabalhos publicados sobre estimação robusta em *componentes principais comuns* porque um dos nossos objectivos é a comparação das nossas propostas com as dos outros autores. No entanto, o principal objectivo deste capítulo é a inclusão de novos resultados robustos. Neste sentido, na Secção 3 deduzimos as *funções de influência parciais* dos funcionais relacionados com os estimadores construídos com as técnicas de “plug-in” e de “projection-pursuit”. A partir destas *funções de influência parciais* derivámos também as variâncias assintóticas dos estimadores relacionados.

## 4.2 Revisão bibliográfica

Nesta subsecção pretendemos apresentar os trabalhos onde foram desenvolvidos e aplicados procedimentos robustos para a estimação das *componentes principais comuns*. Como referimos anteriormente, são poucos os trabalhos publicados nesta área da estatística. De seguida apresentamos por ordem cronológica de publicação os três trabalhos de que temos conhecimento.

**Novi Inverardi, P. L. e Flury, B. N. (1992). Robust estimation of common principal components.**

Neste trabalho os autores alertaram para a necessidade da estimação robusta dos

valores e vectores próprios de matrizes de dispersão que verificam o modelo CPC. Neste sentido, Novi Inverardi e Flury referem a estimação robusta e independente de cada uma das  $k$  matrizes de dispersão, com a posterior substituição das matrizes de covariâncias clássicas de (2.10) e (2.11) por essas matrizes. Foi sugerido o emprego dos estimadores-M de Maronna (1976) para estimar de forma robusta as matrizes de dispersão de cada grupo.

Outra alternativa robusta foi construída generalizando, para  $k$  grupos, a sugestão de Campbell (1980) para o modelo das *componentes principais*. Nesta proposta utiliza-se uma versão robusta do algoritmo F-G, apresentado no Capítulo 2, que os autores referem pela sigla RFGA, consequência das iniciais da designação inglesa “Robust Flury Gautschi Algorithm”. No RFGA é abandonado o princípio da minimização exacta, que caracteriza o núcleo do algoritmo F-G, em prole de um procedimento de estimação com a atribuição de pesos às observações. Os pesos atribuem-se de acordo com a influência que a observação exerce na estimação dos valores próprios da matriz de dispersão. Esta ideia surgiu a Campbell (1980) como consequência das observações poderem ter simultaneamente altas influências numas direcções e baixas noutras. O objectivo deste algoritmo centra-se em encontrar os vectores próprios comuns robustos, e os  $k$  valores próprios robustos associados para cada grupo, que são estimadores-M do modelo CPC. Assim, esta implementação não é mais do que uma aplicação dos estimadores-M de localização e dispersão a cada componente comum de  $\mathbf{S}_1, \dots, \mathbf{S}_k$ , conduzindo às  $p$  direcções comuns sem grande influência das observações discordantes. No primeiro passo do RFGA procede-se à estimação da matriz  $\boldsymbol{\beta}$ , designada por  $\mathbf{B}$ , e das  $k$  matrizes diagonais  $\mathbf{L}_1, \dots, \mathbf{L}_k$  através do algoritmo F-G. Para a determinação da  $m$ -ésima *componente principal comum* robusta constroi-se a projecção das observações na direcção do vector  $\mathbf{b}_m$ , i.e.,

$$\mathbf{u}_{im} = \mathbf{X}_i \mathbf{b}_m, \quad i = 1, \dots, k,$$

com  $\mathbf{X}_i$  a matriz ( $n_i \times p$ ) que contém as observações referentes ao  $i$ -ésimo grupo. De seguida as estimativas-M (localização e dispersão) das variáveis projectadas são sucessivamente utilizadas para ponderar as variáveis originais  $\mathbf{x}_i$ , sempre de acordo com a influência que cada observação exerce nos valores próprios da matriz de covariâncias. Note-se que os pesos atribuídos a cada observação variam de acordo com a componente comum que se está a estimar. Isto é justificado pelo facto já referido de uma observação poder ter influência distinta nos diferentes funcionais. Constroem-se de seguida  $k$  matrizes de covariâncias, a partir dos dados ponderados  $\mathbf{x}_i^*, \mathbf{S}_1^*, \dots, \mathbf{S}_k^*$ , e aplica-se novamente o algoritmo F-G para obter  $\mathbf{B}^*$  e  $\mathbf{L}_1^*, \dots, \mathbf{L}_k^*$ . Quando o critério de convergência estabelecido é alcançado, os  $k$  conjuntos de dados

são projectados no subespaço ortogonal gerado pelos vectores próprios comuns já estimados anteriormente  $\beta_1^*, \dots, \beta_{m-1}^*$ , pois já foi feita a atribuição  $\beta_i^* = \mathbf{b}_i^*$ , para  $i = 1, \dots, m - 1$ . Todo este algoritmo é apresentado com bastante pormenor neste trabalho de Novi Inverardi e Flury. Os autores escolheram a linguagem GAUSS para a implementação do RFGA.

Para avaliar o desempenho desta proposta robusta foi desenvolvido um estudo de Monte Carlo com 2 grupos em dimensão 2, gerando dados de várias distribuições. Consideraram amostras com 5% e 10% de contaminação ( $\varepsilon$ ), da mistura

$$(1 - \varepsilon) N(\mathbf{0}_2, \mathbf{I}_2) + \varepsilon N(\mathbf{0}_2, \sigma^2 \mathbf{I}_2) ,$$

fixando para  $\sigma$  os valores 1, 1.5, 2, 2.5, 3, 4, 5. Além disso, foi utilizado um conjunto de dados reais (Jolicoeur e Mosimann, 1960), também estudados por Flury (1986b) no contexto da análise em *componentes principais comuns*. Neste caso foi introduzida contaminação determinística em três observações de cada um dos dois grupos.

Com os resultados deste estudo foi possível concluir que o desempenho do RFGA para os dados não contaminados é semelhante ao do algoritmo F-G clássico. Em contrapartida, no que refere às situações contaminadas, os resultados da versão robusta RFGA são muito mais resistentes, verificando-se que a influência das observações discordantes não é significativa, contrariamente ao observado com o algoritmo F-G clássico. Assim, os autores concluem que a abordagem desenvolvida e implementada, no seguimento da desenvolvida por Campbell (1980) estima bem os parâmetros do modelo das CPC, protegendo-os contra as contaminações introduzidas. Isto é uma grande vantagem, face à alternativa clássica que facilmente se vê afectada, quer pela não normalidade dos dados, quer pela presença de observações discordantes.

Este trabalho teve o grande mérito de levantar o problema da estimação robusta no tema CPC, e pode eventualmente ter contribuído para o desenvolvimento dos trabalhos subsequentes. No entanto, estes autores não se debruçaram sobre a importante questão da escolha do(s) estimador(es) robusto(s), para a matriz de dispersão, que conduziria a melhores resultados na estimação dos parâmetros do modelo CPC. Esta foi uma questão, entre várias, que ficou em aberto. No entanto, no próximo trabalho desenvolvido nesta área houve a preocupação de abordar esta questão, como veremos de seguida.

**Orellana, L. (1999). Métodos robustos para el análisis de componentes principales comunes.**

Nesta tese de mestrado, a autora centrou a sua investigação na construção de novos métodos robustos para o modelo das CPC e também para o modelo proporcional.

Este último modelo é o Nível 2 do sistema hierárquico de Flury (1988, págs. 60–62), que apresentámos no Capítulo 2.

Considerando estruturas de covariâncias que satisfazem (2.3), as propostas para a estimação robusta dos parâmetros deste modelo basearam-se em:

1. estimar robusta e independentemente as  $k$  matrizes de dispersão, denotadas por  $\mathbf{V}_1, \dots, \mathbf{V}_k$ , e substituir nas equações de verosimilhança (2.10) e (2.11)  $\mathbf{S}_1, \dots, \mathbf{S}_k$  por essas matrizes;
2. obter a matriz combinada  $\mathbf{V} = \frac{1}{N} \sum_{i=1}^k n_i \mathbf{V}_i$  e considerar para estimador de  $\boldsymbol{\beta}$  a matriz  $\hat{\boldsymbol{\beta}}$  construída com os vectores próprios de  $\mathbf{V}$ , enquanto que para estimador da matriz  $\boldsymbol{\Lambda}_i$  se considera a matriz  $\hat{\boldsymbol{\Lambda}}_i = \text{diag}(\hat{\boldsymbol{\beta}}^T \mathbf{V}_i \hat{\boldsymbol{\beta}})$ .

Para o modelo proporcional Orellana propõe uma metodologia análoga à referida em 1, mas considerando as equações de verosimilhança para este caso particular do modelo CPC. Quando se observa a validade de (2.2) as equações de verosimilhança resultantes são,

$$\rho_i = \frac{1}{p} \sum_{j=1}^p \frac{\boldsymbol{\beta}_j^T \mathbf{S}_i \boldsymbol{\beta}_j}{\lambda_j}, \quad i = 2, \dots, k, \quad (4.1)$$

$$\lambda_i = \frac{1}{N - k} \sum_{j=1}^k \frac{(n_i - 1)}{\rho_i} \boldsymbol{\beta}_j^T \mathbf{S}_i \boldsymbol{\beta}_j, \quad j = 1, \dots, p, \quad (4.2)$$

$$\boldsymbol{\beta}_m^T \left( \sum_{i=1}^k \frac{(n_i - 1)}{\rho_i} \mathbf{S}_i \right) \boldsymbol{\beta}_j = 0 \quad m, j = 1, \dots, p, \quad m \neq j, \quad (4.3)$$

com  $N = \sum_{i=1}^k n_i$ , sujeito às restrições de ortogonalidade  $\boldsymbol{\beta}^T \boldsymbol{\beta} = \mathbf{I}_p$ , e com  $\rho_1 = 1$  (Flury, 1988, Capítulo 5).

Para além desta alternativa, foi proposto também o uso de outras duas estatísticas para estimar as constantes de proporcionalidade. Uma destas estatísticas é construída a partir do quociente dos determinantes das matrizes de covariâncias de dois grupos. A outra é a estatística do teste de esfericidade considerada por Muirhead (1982, págs. 333–339).

Foi obtida a distribuição assintótica dos estimadores dos valores próprios, para o modelo CPC nas situações 1 e 2, e também do estimador da constante de proporcionalidade do modelo proporcional.

De forma análoga ao que se observa nas *componentes principais*, no modelo CPC a estimação das matrizes de dispersão é de extrema importância e pode eventualmente afectar todo o processo de estimação dos restantes parâmetros. Aliás, as matrizes de dispersão têm um papel fundamental em qualquer técnica da estatística multivariada. Daí que Orellana tenha reservado um capítulo da sua tese para analisar esta questão, tentando avaliar o comportamento de vários estimadores. Com este objectivo, efectuou um estudo de Monte Carlo para averiguar o desempenho de quatro estimadores robustos das matrizes de dispersão e também do estimador clássico de máxima verosimilhança, com o pressuposto da normalidade. No estudo intervieram dois dos estimadores robustos multivariados que expusemos no Capítulo 3, nomeadamente o estimador-M (Maronna, 1976) e o estimador Stahel-Donoho (Stahel, 1981 e Donoho, 1982). Além destes, também foram seleccionados dois outros estimadores robustos, o estimador de volume elipsoidal mínimo (MVE) de Rousseeuw (1983, 1985) e o estimador-S introduzido inicialmente por Rousseeuw e Yohai (1984) no contexto da análise em regressão. Posteriormente, Davies (1987) generalizou o estimador-S de Rousseeuw e Yohai (1984) para problemas de análise multivariada.

De uma forma grosseira podemos dizer que o MVE procura o elipsóide com volume mínimo que cobre pelo menos  $s$  pontos da amostra. A quantidade  $s$  deve ser escolhida de forma a igualar pelo menos metade do número de observações. Quanto ao estimador-S, este é a solução que minimiza o determinante de  $\mathbf{V}_n$ , com  $\mathbf{V}_n \in PDS(p)$ , sujeita à restrição de

$$n^{-1} \sum_{i=1}^n \rho(d(\mathbf{X}_i, \mathbf{t}_n, \mathbf{V}_n)) = b_0 ,$$

com  $d^2(\mathbf{X}, \mathbf{t}_n, \mathbf{V}_n) = (\mathbf{X} - \mathbf{t}_n)^T \mathbf{V}_n^{-1} (\mathbf{X} - \mathbf{t}_n)$  e  $b_0 > 0$ . Deve-se ainda ter o cuidado de seleccionar  $\rho$  como uma função crescente mas limitada de  $d(\mathbf{X}_i, \mathbf{t}_n, \mathbf{V}_n)$ , Lopuhaä (1989).

Assim, o estimador-S não é mais do que uma versão suave (“smooth”) do MVE, e daí a consequente escolha do nome de estimador-S. Note-se que este último estimador em vez de considerar apenas que determinada observação está dentro ou fora do elipsóide, como faz o MVE, atribui à observação um peso que é função de  $\rho$ . Este estimador pertence à classe dos estimadores multivariados com *alto ponto de ruptura*.

Para avaliar os resultados do seu estudo de simulação Orellana considerou a discrepância entre a matriz estimada e a verdadeira matriz, calculando  $\sum_i \sum_j (v_{ij} - \sigma_{ij})^2$ , onde  $v_{ij}$  é o  $(i, j)$ -ésimo elemento da matriz  $\mathbf{V}$  estimada e  $\sigma_{ij}$  a correspondente en-

trada da verdadeira matriz de dispersão  $\Sigma$ . Além disso, como o interesse se concentra na estimação da forma da matriz  $\Sigma$ , e seguindo o trabalho de Maronna e Yohai (1995), Orellana considerou indicadores da magnitude do enviesamento do estimador  $\mathbf{V}$ , calculando para isso o enviesamento de uma outra matriz  $\mathbf{W} = \left(\Sigma^{-1/2}\right) \mathbf{V} \left(\Sigma^{-1/2}\right)^T$ . Por sua vez, o enviesamento de  $\mathbf{W}$  foi quantificado à custa do desvio em relação à esfericidade desta matriz, usando para isso o quociente entre o maior e o menor valor próprio (excentricidade da matriz). Foi também considerada a estatística do teste de razão de verosimilhanças para avaliar este mesmo desvio.

Com base nos critérios de comparação utilizados, foi possível concluir que o estimador-M e o estimador Stahel-Donoho foram os que se mostraram mais adequados para a estimação das matrizes de dispersão, e portanto a utilizar no modelo das *componentes principais comuns*. Como Orellana realça, ambos os estimadores têm um bom comportamento quando as observações são normalmente distribuídas, sendo os que apresentam menores erros médios e medianos e ainda menores erros quadráticos médios. Deste modo, no estudo de robustez desenvolvido posteriormente a autora utilizou estes dois estimadores robustos.

Foi realizada uma outra simulação de Monte Carlo para comparar o comportamento do estimador de máxima verosimilhança das CPC em relação a propostas resistentes baseadas no estimador-M e no estimador Stahel-Donoho. Os estimadores resistentes considerados neste estudo foram construídos pela técnica referida em 1, para o modelo CPC e proporcional. Consideraram-se duas populações em dimensão 4, com e sem contaminação. As contaminações estudadas abrangeram diversos casos, como por exemplo, contaminações esféricas, simétricas e não simétricas na direcção do vector próprio relacionado com o menor (maior) valor próprio da primeira população.

Em apêndice, a autora apresenta o código dos programas, em linguagem MATLAB, que permitiram desenvolver estes estudos.

A eficácia dos métodos robustos propostos foi também avaliada num conjunto de dados reais (Fisher, 1936), considerando os dados originais, os dados com 4 observações da espécie *Versicolor* modificadas, e ainda o caso onde mais 3 observações da espécie *Virginica* foram alteradas.

Destes estudos, Orellana concluiu que o estimador Stahel-Donoho apresenta um comportamento razoavelmente estável nas contaminações consideradas. O único inconveniente é que produz estimativas enviesadas para os valores próprios. No que se refere ao estimador-M, este exhibe uma maior sensibilidade, e não traduz bons resultados quando existe 10% de contaminação. No entanto, quando o tipo

de contaminação é simétrica este estimador tem um comportamento similar ao do Stahel–Donoho. Por outro lado, uma vantagem do estimador–M é a sua simplicidade computacional e a possibilidade de calibração da eficiência em relação ao estimador de máxima verosimilhança.

Foi ainda possível concluir que a definição de *ponto de ruptura* não é adequada quando se pretende estimar os eixos principais comuns. Como já referimos, o *ponto de ruptura* do funcional matriz de dispersão está relacionado com a “implosão” ou “explosão” da matriz, i.e., o menor valor próprio não tender para zero e o maior valor próprio não tender para infinito. Note-se que nesta definição nada intervém acerca dos vectores próprios da matriz. A autora refere que seria importante a construção de uma definição de *ponto de ruptura* para os funcionais vectores próprios da matriz de dispersão.

Os resultados obtidos neste trabalho, no que diz respeito ao modelo CPC, são também incluídos na publicação Boente e Orellana (2001), que descrevemos seguidamente.

**Boente, G. e Orellana, L. (2001). A robust approach to common principal components.**

As autoras para além de apresentarem os resultados de Orellana (1999), sugeriram uma nova abordagem para a estimação dos parâmetros do modelo CPC. Esta nova técnica é uma generalização da que Li e Chen (1985) propuseram para a análise em *componentes principais*, e baseia-se na técnica de “projection–pursuit”. Este termo poderia eventualmente ser traduzido por *perseguição da projecção*, mas optámos pela designação inglesa. Como o nome sugere, os estimadores de “projection–pursuit” encontram-se através de um processo iterativo que procura a direcção que torna máxima (ou mínima) uma medida de escala robusta dos dados projectados nessa direcção.

Croux e Ruiz–Gazen (1996) desenvolveram um algoritmo rápido para calcular a proposta de Li e Chen (1985), algoritmo esse que as autoras generalizaram para  $k$  populações.

Como exemplo de aplicação, Boente e Orellana escolheram os dados de Fisher (1936) também estudados por Orellana (1999) neste contexto. Considerando os mesmos dados modificados de Orellana (1999), e de forma a resolver os problemas observados na estimação clássica de máxima verosimilhança, foram propostos três métodos robustos de estimação. Obtiveram as estimativas robustas pela técnica 1 de Orellana (1999), atribuindo a  $\mathbf{V}_i$ ,  $i = 1, \dots, k$ , estimadores–M (ME) e estimadores Stahel–

Donoho (SDE), e pela técnica “projection–pursuit” usando o MAD como estimativa robusta de escala (PPE). Além disso, avaliaram ainda as estimativas propostas por Novi Inverardi e Flury (1992), (NFE).

O comportamento da alternativa de NFE e da técnica que utiliza o ME foi semelhante ao dos estimadores de máxima verosimilhança. Bastante melhor foi o método com o SDE, estimando praticamente as mesmas direcções principais comuns dos dados originais. Quanto à técnica PPE, os resultados estão numa situação intermédia, uma vez que os eixos principais dos dados originais foram ligeiramente rodados. Esta foi no entanto a única alternativa que conduziu a resultados satisfatórios para a estimação dos valores próprios.

Foram desenvolvidos dois estudos de Monte Carlo, ambos com duas populações com dimensões 2 e 4. Em ambas as simulações foi observado o comportamento dos seguintes procedimentos: a técnica que utiliza o SDE, a proposta NFE, e duas alternativas “projection–pursuit”, uma construída com o estimador de escala MAD e outra recorrendo a um estimador–M. Em dimensão 4 o desempenho foi ainda comparado com as estimativas de máxima verosimilhança e as obtidas através do ME. Além disso, para  $p = 4$  as autoras ainda obtiveram os eixos principais comuns a partir das matrizes robustas combinadas, método 2 de Orellana (1999). No entanto, os resultados desta abordagem não foram incluídos neste trabalho dada a semelhança com os construídos através da substituição das matrizes clássicas pelas SDE.

Para  $p = 4$ , foram geradas 1000 réplicas de duas amostras independentes com dimensões  $n_1 = n_2 = 100$  e matrizes de dispersão,

$$\Sigma_1 = \text{diag}(1, 2, 3, 4) \quad \text{e} \quad \Sigma_2 = \text{diag}(5, 2.5, 8, 6).$$

Consideraram as três situações distribucionais seguintes:

- $C_0$ : dados normais,  $N(\mathbf{0}_4, \Sigma_i)$ ,  $i = 1, 2$ .
- $C_{1,\varepsilon}$ :  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in}$  vectores aleatórios independentes e identicamente distribuídos (i.i.d.) a  $0.9N(\mathbf{0}_4, \Sigma_i) + 0.1N(\mathbf{0}_4, 9\Sigma_i)$ . Esta contaminação corresponde a inflacionar a matriz de dispersão e assim afecta a variância na direcção dos eixos principais de cada população, mas não a sua orientação.
- $C_{2,\varepsilon}$ :  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in}$  são i.i.d. com  $(1 - \varepsilon)N(\mathbf{0}_4, \Sigma_i) + \varepsilon N(\boldsymbol{\mu}, \Sigma_i)$ , com  $\boldsymbol{\mu} = (10, 0, 0, 0)^T$ . Consideraram as contaminações resultantes de fixar  $\varepsilon = 0.05$  e  $\varepsilon = 0.10$ . Este caso corresponde a contaminar ambas as populações na direcção do mais pequeno valor próprio da primeira população. O objectivo agora é estudar possíveis alterações na estimação das direcções principais.



No que se refere aos valores próprios, as melhores estimativas sob contaminação foram obtidas com o SDE. Quanto ao desempenho do ME, os fracos resultados obtidos foram justificados pelo valor do *ponto de ruptura* empírico, que pode ser muito inferior ao assintótico (Maronna e Yohai, 1995, Adrover, 1998). Neste estudo de simulação o valor do *ponto de ruptura* assintótico é cerca de 13%. Por outro lado, a fraca eficiência das estimativas “projection–pursuit” foi justificada pela eficiência, também fraca, dos estimadores de escala seleccionados. As estimativas dos valores próprios construídas pela proposta de Nori Iverardi e Flury (1992) são as que apresentam o pior desempenho sob a contaminação  $C_{1,\varepsilon}$ . Nas restantes situações o comportamento do NFE é similar ao do SDE.

Para avaliar os resultados do processo de estimação das direcções principais comuns, foi calculado o quadrado da distância entre a verdadeira direcção principal (vector da base canónica) e a direcção principal estimada. Quando os dados são afectados pela contaminação assimétrica  $C_{2,\varepsilon}$  os estimadores–M têm um comportamento bastante fraco, comparável ao dos estimadores de máxima verosimilhança, enquanto que o SDE e o NFE são muito menos sensíveis a este tipo de contaminação. As propostas “projection–pursuit” também se mostraram sensíveis à contaminação imposta. Isto pode ser justificado pelo facto da matriz  $\Sigma_{\text{id}} = \sum_{i=1}^k \tau_i \Sigma_i = \text{diag}(3, 2.25, 5.5, 5)$  ter valores próprios próximos. As dimensões das amostras utilizadas poderão não ser suficientes para distinguir claramente as duas primeiras e as duas últimas direcções principais comuns. Posteriormente, na Subsecção 3.2 deste capítulo, será justificada a necessidade da construção da matriz  $\Sigma_{\text{id}}$ . Para a alternativa 2 de Orellana (1999), as autoras referem o mesmo problema, pois esta hipótese também utiliza a matriz combinada. Em conclusão, todas as alternativas se mostraram bastante sensíveis à contaminação quando o objectivo da estimação são as direcções principais.

No estudo de Monte Carlo realizado em dimensão 2, tomou-se  $n_1 = n_2 = n = 50$  e considerou-se apenas 500 replicações. A redução do número de replicações para metade justifica-se pelos métodos de “projection–pursuit” e o SDE utilizarem uma grelha suficientemente fina (1000 direcções), em vez de recorrer a técnicas de reamostragem como é necessário em dimensões superiores. Este estudo também foi delineado de forma a evitar o problema resultante da proximidade dos valores próprios da matriz combinada. Neste sentido, as duas matrizes de dispersão teóricas escolhidas foram,

$$\Sigma_1 = \text{diag}(4, 14) \quad \text{e} \quad \Sigma_2 = \text{diag}(2, 12),$$

que conduzem a  $\Sigma_{\text{id}} = \text{diag}(3, 13)$ .

As situações distribucionais consideradas foram,

- $C_0$ : dados normais,  $N(\mathbf{0}_2, \mathbf{\Sigma}_i), i = 1, 2$ .
- $C_{1,\varepsilon}$ :  $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n}$  são vectores aleatórios i.i.d. a  $0.8N(\mathbf{0}_2, \mathbf{\Sigma}_1) + 0.1\Delta\boldsymbol{\mu} + 0.1\Delta_{-\boldsymbol{\mu}}$  com  $\boldsymbol{\mu} = (4.72; -0.042)^T$  e  $\mathbf{X}_{21}, \dots, \mathbf{X}_{2n}$  i.i.d. com distribuição não contaminada  $N(\mathbf{0}_2, \mathbf{\Sigma}_2)$ .

Deste processo de simulação, Boente e Orellana concluíram que o comportamento das alternativas via SDE e NFE é semelhante ao observado em dimensão 4. Por outro lado, as propostas “projection–pursuit” são agora muito mais resistentes, salientando-se favoravelmente a alternativa com o MAD. Conclui-se que estas novas propostas “projection–pursuit” são uma boa ferramenta para a estimação dos eixos principais comuns.

Foi também realçado por Boente e Orellana que a noção de *ponto de ruptura* não é apropriada para a estimação das direcções principais. Neste estudo foi evidente a sensibilidade dos estimadores dos vectores próprios face à contaminação, embora o valor máximo fixado para a contaminação fosse 20% e muito inferior ao *ponto de ruptura*. As autoras sugerem uma noção de *ponto de ruptura* que tenha em conta o coseno do ângulo entre a verdadeira direcção e a estimada. Orellana (1999) já tinha alertado para esta questão.

### 4.3 Novos resultados

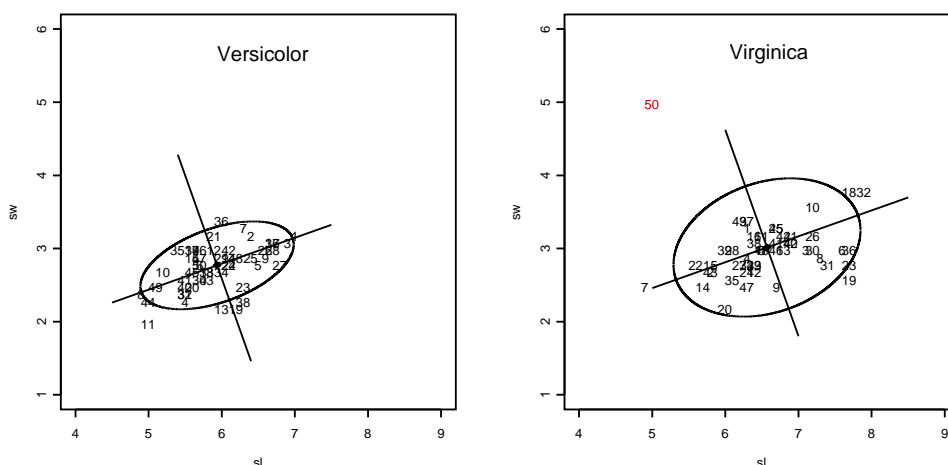
Como se observou na secção anterior, em nenhum dos três trabalhos expostos foram deduzidas as *funções de influência* para os funcionais em causa. Como foi referido anteriormente, foram vários os autores que noutros contextos utilizaram estas funções para derivar novas propostas robustas, em especial para a detecção de observações discordantes. Achamos assim importante dispor desta metodologia robusta para a análise em *componentes principais comuns*. De notar que para o modelo em estudo, que trata de funcionais relacionados com  $k$  populações simultaneamente, será necessário uma generalização da *função de influência*.

No Capítulo 1 ilustrámos a necessidade de procedimentos robustos para a estimação das *componentes principais*. De modo semelhante ao caso da análise em *componentes principais*, a generalização para  $k$  grupos também pode ser afectada quer pela presença de observações discordantes, quer por desvios à distribuição normal. Relembre-se que a não normalidade dos dados conduz à perda de algumas das “boas” propriedades do estimador clássico da matriz de covariâncias,  $\mathbf{S}_i$ .

Na tentativa de ilustrar a necessidade de procedimentos robustos na estimação das CPC, considerámos o exemplo de Fisher (1936) referido no Capítulo 1, com a observação 50 do grupo *Virginica* modificada. Aplicámos de novo o programa de Phillips (1994) para estimar os vectores próprios comuns, assim como as variâncias associadas em cada grupo. Os resultados encontrados indicam-se na Tabela 4.1.

	vectores próprios			valores próprios	
<i>sl</i>	0.943	-0.334	<i>Versicolor</i>	0.301	0.064
<i>sw</i>	0.334	0.942	<i>Virginica</i>	0.435	0.197

**Tabela 4.1:** Análise em CPC.



**Figura 4.1:** Elipses correspondentes a distância de Mahalanobis constante, sob CPC, e eixos das CPC.

Construímos também os novos contornos com distância de Mahalanobis constante,  $(\mathbf{x} - \bar{\mathbf{x}}_i)^T \hat{\Sigma}_i (\mathbf{x} - \bar{\mathbf{x}}_i) = 4$ . Na Figura 4.1 apresentam-se estes contornos com a sobreposição dos eixos principais comuns.

Para avaliar a influência, ou não, das observações discordantes na estimação dos parâmetros do modelo CPC, procedemos também à estimação robusta das *componentes principais comuns* e respectivas variâncias. A técnica robusta utilizada foi a mais intuitiva possível. Esta consiste, em proceder primeiro à estimação robusta das matrizes de dispersão de cada grupo (independentemente). De seguida, substituir nas equações de verosimilhança (2.10) e (2.11) as matrizes  $\mathbf{S}_i$  pelas matrizes

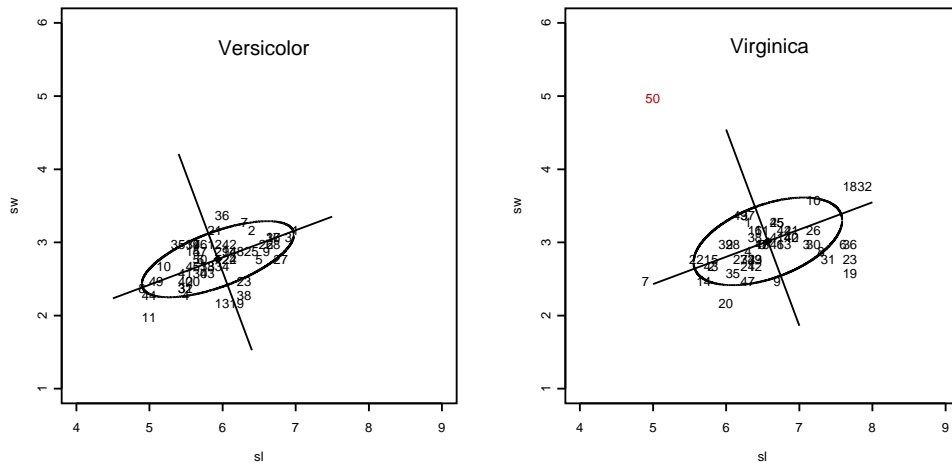
de dispersão robustas encontradas. Resolvendo este sistema de equações, sujeito à condição (2.12), encontram-se umas estimativas robustas dos vectores próprios comuns e dos valores próprios associados a cada grupo. Note-se que a solução robusta encontrada está dependente da escolha dos estimadores robustos considerados para as  $k$  matrizes de dispersão. Este procedimento é referido na literatura pela designação “plug-in” e, como vimos na Secção 2, foi considerado pelos autores Novi Inverardi e Flury (1992), Orellana (1999) e Boente e Orellana (2001).

A técnica “plug-in” foi aplicada com o estimador RMCD, apresentado no Capítulo 3, Secção 8, fixando o subconjunto dos dados a utilizar em  $h = [0.75n]$ . Os resultados desta estimação apresentam-se na Tabela 4.2. Foram ainda feitos gráficos seme-

	vectores próprios			valores próprios	
<i>sl</i>	0.937	-0.349	<i>Versicolor</i>	0.303	0.035
<i>sw</i>	0.349	0.937	<i>Virginica</i>	0.285	0.063

**Tabela 4.2:** Análise robusta em CPC.

lhantes aos da Figura 4.1, mas considerando uma versão robusta da distância de Mahalanobis. Seguindo o trabalho de Rousseeuw e van Zomeren (1990), calculámos os contornos limitados por  $(\mathbf{x} - \hat{\boldsymbol{\mu}}_i)^T \mathbf{V}_i (\mathbf{x} - \hat{\boldsymbol{\mu}}_i) = 4$ , com  $\mathbf{V}_i$  e  $\hat{\boldsymbol{\mu}}_i$ , respectivamente, as estimativas RMCD de dispersão e localização. Na Figura 4.2 encontram-se esses



**Figura 4.2:** Elipses correspondentes a distância de Mahalanobis robusta constante, sob CPC, e eixos das CPC robustas.

contornos com a sobreposição dos eixos das *componentes principais comuns* robustas.

Como se pode constatar pelos resultados obtidos, as direcções das *componentes principais comuns* clássicas e robustas são quase paralelas. Parece assim ser possível concluir que os eixos principais comuns não sofreram alterações significativas com a modificação da última observação do grupo *Virginica*. Por outro lado, nota-se com a estimação robusta das CPC uma redução significativa nos valores próprios. Nota-se ainda que, de modo semelhante ao caso das CP apresentado na Figura 1.2, os contornos de densidade constante no grupo *Virginica* da Figura 4.1 aumentaram devido ao novo valor da observação 50. É fácil perceber que o aumento do número, ou do valor, das observações discordantes poderia facilmente conduzir a estimação clássica a resultados completamente inadequados.

Em conclusão, é importante dispor-se de procedimentos robustos para a estimação das *componentes principais comuns*. Para além disso, é desejável a comparação do grau de robustez dos diferentes procedimentos.

Como veremos nas Subsecções 3.1 e 3.2 deste capítulo, para a estimação robusta das *componentes principais comuns* é sempre necessário considerar estimadores de dispersão robustos, univariados ou multivariados. Assim, uma questão que se coloca é qual será o estimador robusto mais adequado. Como se sabe, é sempre preferível considerar um estimador com *alto ponto de ruptura*. Mas, dentro desta classe de estimadores, qual será o estimador mais apropriado?

Com o objectivo de esclarecer esta questão, recentemente Croux e Haesbroeck (2000) estudaram este problema no caso de uma população, i.e., modelo das *componentes principais*. O trabalho destes autores foi baseado em estudos de simulação de Monte Carlo e na *função de influência* de alguns funcionais. Como salientámos no Capítulo 3, a forma da *função de influência* permite avaliar a robustez do estimador relacionado (equivalente) com o funcional em causa. Croux e Haesbroeck (2000) indicaram qual o estimador, dentro dos utilizados, mais apropriado para a estimação robusta das *componentes principais*. Além disso, outros autores como Critchley (1985) e Shi (1997) utilizaram a *função de influência*, no contexto da análise em *componentes principais*, para a construção de métodos gráficos de diagnóstico e detecção de observações influentes.

Assim, é nosso objectivo, em primeiro lugar, calcular a *função de influência* para o modelo das *componentes principais comuns*. Como já referimos, no modelo das *componentes principais comuns* tratamos com  $k$  grupos em simultâneo e a *função de influência* de Hampel (1968, 1974) foi definida para problemas com funcionais num só grupo. É então necessária uma generalização da *função de influência* para problemas

com funcionais dependentes de várias populações, e onde se admite que apenas a  $i$ -ésima população sofre contaminação. Embora este problema já tivesse sido focado por outros autores, como por exemplo Rousseeuw e Ronchetti (1981) e Hampel *et al.* (1986), foram Pires (1995) e Pires e Branco (2002) que apresentaram uma definição explícita para esta generalização da *função de influência*, a que chamaram *funções de influência parciais*.

Como os funcionais estudados neste trabalho são  $p$ -variados apresenta-se a definição das *funções de influência parciais* para esta situação.

**Definição 4.1 (funções de influência parciais  $p$ -dimensionais, Pires, 1995, Pires e Branco, 2002)** Seja  $F$  a medida produto,  $F = F_1 \times \cdots \times F_k$ . Considere-se  $F_{\mathbf{x}, \varepsilon, i_0} = F_1 \times \cdots \times F_{i_0-1} \times F_{i_0, \mathbf{x}, \varepsilon} \times F_{i_0+1} \times \cdots \times F_k$  com  $F_{i, \mathbf{x}, \varepsilon} = (1 - \varepsilon)F_i + \varepsilon\Delta_{\mathbf{x}}$ .

As *funções de influência parciais* (PIF) do funcional  $\mathbf{T}$ ,  $p$ -dimensional, em  $F$  são dadas por

$$\text{PIF}_{i_0}(\mathbf{x}, \mathbf{T}, F) = \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{T}(F_{\mathbf{x}, \varepsilon, i_0}) - \mathbf{T}(F)}{\varepsilon}, \quad i_0 = 1, \dots, k, \quad (4.4)$$

com  $\mathbf{x} \in \Omega$  e para o qual este limite existe.

Foi a analogia com a noção de derivada parcial que levou os autores referidos anteriormente a designarem estas funções por *funções de influência parciais*, como se pode comprovar com a definição que se segue.

**Definição 4.2 (diferenciação parcial à Gâteaux)** Seja  $F$  a medida produto,  $F = F_1 \times \cdots \times F_k$ . Considere-se  $GE(\varepsilon, F_{i_0}) = F_1 \times \cdots \times F_{i_0-1} \times GE(F_{i_0}, \varepsilon) \times F_{i_0+1} \times \cdots \times F_k$  com  $GE(F_i, \varepsilon) = (1 - \varepsilon)F_i + \varepsilon H_i$ . O funcional  $\mathbf{T}$ ,  $p$ -dimensional, diz-se parcialmente diferenciável à Gâteaux em  $F$ , em ordem a  $F_{i_0}$ , se existir uma função  $a_{i_0}(\mathbf{x})$  tal que

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbf{T}[GE(\varepsilon, F_{i_0})] - \mathbf{T}(F)}{\varepsilon} = \int a_{i_0}(\mathbf{x}) dH_{i_0}(\mathbf{x}).$$

Comparando estas duas definições é fácil concluir que no caso do funcional  $\mathbf{T}$  ser parcialmente diferenciável à Gâteaux em  $F$  em ordem a  $F_{i_0}$  então

$$a_{i_0}(\mathbf{x}) = \text{PIF}_{i_0}(\mathbf{x}, \mathbf{T}, F),$$

e além disso

$$\int a_{i_0}(\mathbf{x}) dF_{i_0}(\mathbf{x}) = \int \text{PIF}_{i_0}(\mathbf{x}, \mathbf{T}, F) dF_{i_0}(\mathbf{x}) = \mathbf{0}_p.$$

Pires e Branco (2002) mostraram também que o resultado apresentado em (3.20), no caso do funcional  $\mathbf{T}$  depender de  $k$  populações, generaliza-se para

$$N^{\frac{1}{2}} \{ \mathbf{T}(F_N) - \mathbf{T}(F) \} = \sum_{i=1}^k \frac{1}{(\tau_i n_i)^{\frac{1}{2}}} \sum_{j=1}^{n_i} \text{PIF}_i(\mathbf{x}_{ij}, \mathbf{T}, F) + \sqrt{N} R(F_N - F) , \quad (4.5)$$

onde  $F_N$  representa a função de distribuição empírica de  $k$  amostras independentes  $\mathbf{X}_{ij}$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n_i$ .

Quando  $\sqrt{N} R(F_N - F) \xrightarrow{p} 0$ , então pelo teorema do limite central e lema de Slutsky resulta que,  $N^{\frac{1}{2}} [\mathbf{T}(F_N) - \mathbf{T}(F)]$  tem distribuição assymptótica normal  $p$ -variada com média  $\mathbf{0}_p$  e matriz de covariâncias assymptótica dada por

$$\text{ASVAR}(\mathbf{T}, F) = \sum_{i=1}^k \tau_i^{-1} E_{F_i} \left\{ \text{PIF}_i(\mathbf{X}_{i1}, \mathbf{T}, F) \text{PIF}_i(\mathbf{X}_{i1}, \mathbf{T}, F)^T \right\} . \quad (4.6)$$

O resultado (4.6) é bastante útil para a dedução da variância assymptótica do estimador equivalente ao funcional  $\mathbf{T}$ .

A partir das *funções de influência parciais* é possível definir medidas sumárias equivalentes às introduzidas por Hampel (1968). Uma outra noção, também bastante importante na área da estatística robusta e introduzida por Rousseeuw (1981), é a de *B-robustez*. Quando os estimadores se referem a parâmetros que dependem de várias distribuições o conceito de *B-robustez* generaliza-se da forma que se segue.

**Definição 4.3 (estimador B-robusto com dependência de várias populações, Pires 1995)** Um estimador equivalente a um funcional  $\mathbf{T}$  que dependa de várias distribuições é *B-robusto* se todas as suas *funções de influência parciais* são limitadas, i.e,

$$\gamma_i^* = \gamma_i^*(\mathbf{T}, F) = \sup \text{PIF}_i(\mathbf{x}, \mathbf{T}, F) < \infty , \quad i = 1, \dots, k .$$

Como exemplo, consideramos  $F_i$  a função de distribuição de  $\mathbf{X}_i$ ,  $i = 1, \dots, k$ , vector aleatório de  $\mathbb{R}^p$  com vector média  $\boldsymbol{\mu}_i$  e matriz de dispersão  $\boldsymbol{\Sigma}_i \in PDS(p)$  e  $\mathbf{S}_i$  o estimador clássico usual e não corrigido da matriz de covariâncias  $\boldsymbol{\Sigma}_i$  definido por  $\mathbf{S}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i) (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^T$ , com  $\bar{\mathbf{X}}_i = \sum_{j=1}^{n_i} \frac{\mathbf{X}_{ij}}{n_i}$  e  $i = 1, \dots, k$ . Note-se que é necessário utilizar o estimador não centrado para garantir a equivalência entre estimador e funcional da matriz de covariâncias<sup>1</sup>. No entanto, o resultado

---

<sup>1</sup>Ver resultado da pág. 46.

que apresentamos de seguida é aproximadamente válido quando  $\mathbf{S}_i$  é o estimador centrado da matriz de covariâncias da  $i$ -ésima população.

Considere-se  $\mathbf{V}_i(F_i) = \mathbf{V}_i$  o funcional da matriz de covariâncias da  $i$ -ésima população equivalente ao estimador  $\mathbf{S}_i$ . De forma análoga ao resultado da página 56, a *função de influência* deste funcional vem

$$\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) = (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^T - \boldsymbol{\Sigma}_i. \quad (4.7)$$

Admita-se agora que consideramos o caso em que as  $k$  populações possuem a mesma matriz de covariâncias, ou seja (2.1). Um estimador clássico natural para  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_k$  é a matriz de covariâncias amostral combinada,  $\mathbf{S}_P = \sum_{i=1}^k \tau_i \mathbf{S}_i$ . Como consequência, o estimador  $\mathbf{S}_P$  refere-se a um parâmetro,  $\boldsymbol{\Sigma}$ , que depende de  $k$  populações e o estudo das suas propriedades de robustez não deve descurar esta situação.

Nesse sentido, considera-se  $F$  a distribuição resultante da medida produto,  $F = F_1 \times \dots \times F_k$  e  $\mathbf{V}(F) = \mathbf{V}$  o funcional equivalente ao estimador  $\mathbf{S}_P$ . Qual será a alteração sofrida no funcional  $\mathbf{V}$  com a colocação de uma massa pontual no ponto  $\mathbf{x}_i$  com probabilidade  $\varepsilon$  no  $i$ -ésimo grupo? A resposta a esta questão é dada pela *função de influência parcial* da  $i$ -ésima população.

É fácil verificar que para o funcional  $\mathbf{V}$  as *funções de influência parciais* são

$$\text{PIF}_i(\mathbf{x}, \mathbf{V}, F) = \tau_i \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) = \tau_i \left[ (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^T - \boldsymbol{\Sigma} \right], \quad (4.8)$$

o que evidencia a falta de *B-robustez* do estimador  $\mathbf{S}_P$ .

Estes novos conceitos de robustez infinitesimal permitiram-nos construir as *funções de influência parciais* para os funcionais relacionados com os estimadores do modelo das *componentes principais comuns* e, com isso, propor novas alternativas robustas nesta área da estatística.

Foram deduzidas as *funções de influência parciais* dos funcionais relacionados com os estimadores obtidos com as duas alternativas seguintes:

- técnica de “plug-in”;
- técnica de “projection-pursuit”.

Enquanto decorria o nosso trabalho, em simultâneo, independentemente e sem ser do nosso conhecimento a Professora Graciela Boente desenvolvia, na Argentina, uma investigação análoga, também com o estudo infinitesimal das características de



robustez dos funcionais do modelo CPC, e obviamente com a construção das *funções de influência parciais*. Quando confrontados os resultados de ambas as investigações, resolveu-se efectuar a publicação conjunta Boente, Pires e Rodrigues (2002a).

### 4.3.1 Funções de influência parciais e variâncias assintóticas dos estimadores “plug-in”

Nesta subsecção vamos apresentar a dedução das *funções de influência parciais* dos funcionais relacionados com os estimadores “plug-in”. Para isso, considerámos um sistema de equações semelhante ao de máxima verosimilhança de Flury (1984), apresentado em (2.10), (2.11) e (2.12), mas com  $n_i$  em vez de  $(n_i - 1)$  para garantir a equivalência entre funcionais e estimadores, substituindo as matrizes de covariâncias amostrais por matrizes de dispersão robustas. Os estimadores “plug-in” vão ser então definidos como a solução de

$$\text{diag}(\boldsymbol{\beta}^T \mathbf{V}_i \boldsymbol{\beta}) = \boldsymbol{\Lambda}_i, \quad (4.9)$$

$$\boldsymbol{\beta}_m^T \left( \sum_{i=1}^k n_i \frac{\lambda_{im} - \lambda_{ij}}{\widehat{\lambda}_{im} \lambda_{ij}} \mathbf{V}_i \right) \boldsymbol{\beta}_j = 0, \quad \text{para } m \neq j, \quad (4.10)$$

$$\boldsymbol{\beta}_m^T \boldsymbol{\beta}_j = \delta_{mj}, \quad (4.11)$$

onde  $\mathbf{V}_i$  é uma matriz de dispersão robusta e consistente para  $\boldsymbol{\Sigma}_i$  e  $\delta_{mj}$  é o símbolo de Kronecker.

Para a construção das *funções de influência parciais*, considera-se uma dada distribuição  $F$  definida pela medida produto  $F = F_1 \times \cdots \times F_k$ , com  $\mathbf{X}_i \sim F_i$ , sendo  $F_i$  uma distribuição com parâmetro de localização  $\boldsymbol{\mu}_i$  e matriz de dispersão  $\boldsymbol{\Sigma}_i$  que satisfaz o modelo (2.3). De forma a simplificar os cálculos inerentes à derivação das *funções de influência parciais*, e sem qualquer perda de generalidade, assumimos que os  $k$  vectores média são o vector nulo de  $\mathbb{R}^p$ . Representa-se por  $\mathbf{V}_i = \mathbf{V}_i(F_i)$  o funcional da matriz de dispersão avaliado na distribuição do  $i$ -ésimo grupo. Considera-se ainda o par de funcionais  $\boldsymbol{\beta}_{\mathbf{V}}(F)$ ,  $\boldsymbol{\Lambda}_{\mathbf{V},i}(F)$ ,  $i = 1, \dots, k$ , com  $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_k)$ , que é solução do sistema

$$\text{diag} \left\{ \boldsymbol{\beta}_{\mathbf{V}}(F)^T \mathbf{V}_i(F_i) \boldsymbol{\beta}_{\mathbf{V}}(F) \right\} = \boldsymbol{\Lambda}_{\mathbf{V},i}(F), \quad (4.12)$$

$$\boldsymbol{\beta}_{\mathbf{V},m}(F)^T \left\{ \sum_{i=1}^k \tau_i \frac{\lambda_{\mathbf{V},im}(F) - \lambda_{\mathbf{V},ij}(F)}{\lambda_{\mathbf{V},im}(F) \lambda_{\mathbf{V},ij}(F)} \mathbf{V}_i(F_i) \right\} \boldsymbol{\beta}_{\mathbf{V},j}(F) = 0, \quad m \neq j, \quad (4.13)$$

$$\boldsymbol{\beta}_{\mathbf{V},m}(F)^T \boldsymbol{\beta}_{\mathbf{V},j}(F) = \delta_{mj}. \quad (4.14)$$

Nos casos em que o funcional  $\mathbf{V}_i$  é Fisher-consistente para o parâmetro  $\Sigma_i$ , i.e. quando se verifica a igualdade  $\mathbf{V}_i(F_i) = \Sigma_i$ , então as soluções  $(\Lambda_{\mathbf{V},i}(F), \beta_{\mathbf{V}}(F))$  são Fisher-consistentes para os parâmetros  $(\Lambda_i, \beta)$  do modelo das *componentes principais comuns*.

Além disso, quando o funcional  $\mathbf{V}_i = \mathbf{V}_i(F_i)$  é *equivariante para transformações afins*, ou seja

$$\mathbf{V}_i(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbf{V}_i(\mathbf{X})\mathbf{A}^T,$$

os funcionais  $\beta_{\mathbf{V},m}$  e  $\lambda_{\mathbf{V},im}$  possuem *equivariância ortogonal*. A demonstração deste resultado encontra-se na proposição que se segue.

**Proposição 4.1 (equivariância ortogonal dos funcionais das CPC)**

Seja  $\mathbf{X}_i$ ,  $i = 1, \dots, k$ , um vector de  $\mathbb{R}^p$  com distribuição  $F_i$  e matriz de dispersão  $\Sigma_i$  que satisfaz (2.3). Represente-se por  $\mathbf{V}_i(F_i)$  o funcional da matriz de dispersão do  $i$ -ésimo grupo, *equivariante face a transformações afins*. Sejam  $\Lambda_{\mathbf{V},i}(F)$  e  $\beta_{\mathbf{V}}(F)$  os funcionais solução do sistema (4.12), (4.13) e (4.14). Considerando  $\mathbf{U}_i = \Gamma\mathbf{X}_i$ , com  $\Gamma$  uma matriz ortogonal ( $p \times p$ ), então os funcionais solução do sistema rodado ortogonalmente,  $(\Lambda_{\mathbf{V},i}^*(G), \beta_{\mathbf{V}}^*(G))$ , verificam  $\beta_{\mathbf{V},m}(F) = \Gamma\beta_{\mathbf{V},m}^*(G)$  e  $\lambda_{\mathbf{V},im}(F) = \lambda_{\mathbf{V},im}^*(G)$ , onde  $G = G_1 \times \dots \times G_k$  é a distribuição de  $\mathbf{U}_i$ .

**Dem:** Como o funcional  $\mathbf{V}_i(F_i)$  é *equivariante para transformações afins* resulta que  $\mathbf{V}_{\mathbf{U}_i}(G_i) = \Gamma\mathbf{V}_i(F_i)\Gamma^T$ , e da ortogonalidade de  $\Gamma$  vem que  $\mathbf{V}_i(F_i) = \Gamma^T\mathbf{V}_{\mathbf{U}_i}(G_i)\Gamma$ . Substituindo no sistema (4.12), (4.13) e (4.14) tem-se

$$\begin{aligned} \text{diag} \left\{ \beta_{\mathbf{V}}(G)^T \Gamma^T \mathbf{V}_{\mathbf{U}_i}(G_i) \Gamma \beta_{\mathbf{V}}(G) \right\} &= \Lambda_{\mathbf{V},i}(G), \\ \beta_{\mathbf{V},m}(G)^T \Gamma^T \left\{ \sum_{i=1}^k \tau_i \frac{\lambda_{\mathbf{V},im}(G) - \lambda_{\mathbf{V},ij}(G)}{\lambda_{\mathbf{V},im}(G)\lambda_{\mathbf{V},ij}(G)} \mathbf{V}_{\mathbf{U}_i}(G_i) \right\} \Gamma \beta_{\mathbf{V},j}(G) &= 0, \quad m \neq j, \\ \beta_{\mathbf{V}}(G)^T \Gamma^T \Gamma \beta_{\mathbf{V}}(G) &= \mathbf{I}_p, \end{aligned}$$

implicando as igualdades que  $\beta_{\mathbf{V}}^*(G) = \Gamma\beta_{\mathbf{V}}(F)$  e  $\Lambda_{\mathbf{V},i}^*(G) = \Lambda_{\mathbf{V},i}(F)$ .  $\square$

No teorema seguinte apresentamos as *funções de influência parciais* dos funcionais equivalentes aos estimadores “plug-in” dos parâmetros do modelo das CPC.

**Teorema 4.1 (funções de influência parciais dos funcionais equivalentes aos estimadores “plug-in” do modelo CPC)**

Seja  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, \dots, k$ , uma amostra aleatória da distribuição  $F_i$ , com parâmetro de localização  $\mu_i = \mathbf{0}_p$  e matriz de dispersão  $\Sigma_i$  que satisfaz o modelo (2.3). Represente-se por  $F$  a distribuição definida pela medida produto  $F =$

$F_1 \times \cdots \times F_k$ , e assumamos que,  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e  $\sum_{i=1}^k \tau_i = 1$ .

Considere-se  $\mathbf{V}_i(F_i)$  um funcional consistente segundo Fisher da matriz de dispersão, i.e.,  $\mathbf{V}_i(F_i) = \Sigma_i$ . Represente-se por  $(\lambda_{i1}, \dots, \lambda_{ip})$  e  $(\beta_1, \dots, \beta_p)$ , respectivamente os valores próprios e os vectores próprios comuns da matriz  $\Sigma_i$ . Assuma-se ainda que  $\lambda_{11} > \cdots > \lambda_{1p}$  e que  $\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i)$  existe.

Então, as *funções de influência parciais* das soluções  $(\Lambda_{\mathbf{V},i}(F), \beta_{\mathbf{V}}(F))$  de (4.12) a (4.14) são dadas por

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V}, \ell j}, F) = \delta_{\ell i} \beta_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \beta_j, \quad (4.15)$$

$$\text{PIF}_i(\mathbf{x}, \beta_{\mathbf{V}, j}, F) = \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im} \lambda_{ij}} \left\{ \sum_{\ell=1}^k \frac{\tau_\ell (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right\}^{-1} \{ \beta_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \beta_m \} \beta_m. \quad (4.16)$$

**Dem:** A técnica que utilizámos nesta demonstração é semelhante à usada por Croux e Haesbroeck (2000) na demonstração do Lema 3.

Considere-se a seguinte mistura  $F_{i,\mathbf{x},\varepsilon} = (1 - \varepsilon)F_i + \varepsilon\Delta_{\mathbf{x}}$  e seja  $F_{\mathbf{x},\varepsilon,i} = F_1 \times \cdots \times F_{i-1} \times F_{i,\mathbf{x},\varepsilon} \times F_{i+1} \times \cdots \times F_k$  a medida produto com contaminação na  $i$ -ésima população. Considerem-se também os seguintes funcionais sob contaminação:

- $\beta_{j,\varepsilon,i} = \beta_{\mathbf{V},j}(F_{\mathbf{x},\varepsilon,i})$ , o funcional relacionado com o estimador do  $j$ -ésimo vector próprio e com a  $i$ -ésima população sofrendo contaminação;
- $\lambda_{\ell j,\varepsilon,i} = \lambda_{\mathbf{V},\ell j}(F_{\mathbf{x},\varepsilon,i})$  o funcional relacionado com o estimador do  $\ell$ -ésimo valor próprio da população  $j$  com a  $i$ -ésima população contaminada pontualmente em  $\mathbf{x}$ ;
- $\mathbf{V}_{i,\varepsilon} = \mathbf{V}_i(F_{i,\mathbf{x},\varepsilon})$  o funcional sob a contaminação da  $i$ -ésima população, equivalente ao estimador da matriz de dispersão da  $i$ -ésima população.

Represente-se por  $\mathbf{V}_\ell = \mathbf{V}_\ell(F_\ell)$  o funcional equivalente ao estimador da matriz de dispersão da  $\ell$ -ésima população, sem qualquer população contaminada.

As *funções de influência parciais* das soluções  $(\Lambda_{\mathbf{V},i}(F), \beta_{\mathbf{V}}(F))$  de (4.12) a (4.14) podem ser construídas substituindo-se nas equações (4.12), (4.13) e (4.14)  $F$  por  $F_{\mathbf{x},\varepsilon,i}$  e posteriormente avaliando a sua derivada em ordem a  $\varepsilon$ .

Assim, com a introdução da contaminação na  $i$ -ésima população resulta o sistema

$$\lambda_{ij,\varepsilon,i} = \boldsymbol{\beta}_{j,\varepsilon,i}^T \mathbf{V}_{i,\varepsilon} \boldsymbol{\beta}_{j,\varepsilon,i} , \quad (4.17)$$

$$\lambda_{\ell j,\varepsilon,i} = \boldsymbol{\beta}_{j,\varepsilon,i}^T \mathbf{V}_{\ell} \boldsymbol{\beta}_{j,\varepsilon,i} , \quad \text{para } \ell \neq i , \quad (4.18)$$

$$\begin{aligned} 0 &= \boldsymbol{\beta}_{m,\varepsilon,i}^T \left( \sum_{\ell=1, \ell \neq i}^k \tau_{\ell} \frac{\lambda_{\ell m,\varepsilon,i} - \lambda_{\ell j,\varepsilon,i}}{\lambda_{\ell m,\varepsilon,i} \lambda_{\ell j,\varepsilon,i}} \mathbf{V}_{\ell} \right) \boldsymbol{\beta}_{j,\varepsilon,i} + \\ &+ \boldsymbol{\beta}_{m,\varepsilon,i}^T \left( \tau_i \frac{\lambda_{im,\varepsilon,i} - \lambda_{ij,\varepsilon,i}}{\lambda_{im,\varepsilon,i} \lambda_{ij,\varepsilon,i}} \mathbf{V}_{i,\varepsilon} \right) \boldsymbol{\beta}_{j,\varepsilon,i} , \quad \text{para } m \neq j , \end{aligned} \quad (4.19)$$

$$\delta_{mj} = \boldsymbol{\beta}_{m,\varepsilon,i}^T \boldsymbol{\beta}_{j,\varepsilon,i} . \quad (4.20)$$

• *Funções de influência parciais dos valores próprios.*

Derivando a equação (4.20) em ordem a  $\varepsilon$ , obtém-se

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},m}, F)^T \boldsymbol{\beta}_m = 0 , \quad \text{para } m = j , \quad (4.21)$$

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},m}, F)^T \boldsymbol{\beta}_j + \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_m = 0 , \quad \text{para } m \neq j . \quad (4.22)$$

Com a diferenciação da equação (4.17) resulta

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F) = \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_i \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) .$$

Usando o facto de  $\boldsymbol{\Sigma}_{\ell} \boldsymbol{\beta}_j = \lambda_{\ell j} \boldsymbol{\beta}_j$ , obtém-se

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},ij}, F) = 2\lambda_{ij} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j , \quad (4.23)$$

ficando com a aplicação de (4.21) a igualdade

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},ij}, F) = \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j , \quad (4.24)$$

obtendo-se assim (4.15) para  $\ell = i$ . Por outro lado, derivando em ordem a  $\varepsilon$  a equação (4.18) resulta

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F) = \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\Sigma}_{\ell} \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_{\ell} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) .$$

Usando novamente a relação  $\boldsymbol{\Sigma}_{\ell} \boldsymbol{\beta}_j = \lambda_{\ell j} \boldsymbol{\beta}_j$  vem

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F) = 2\lambda_{\ell j} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_j . \quad (4.25)$$

Aplicando a (4.25) a condição de ortogonalidade (4.21) obtém-se

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F) = 0 , \quad \text{para } \ell \neq i , \quad (4.26)$$

o que implica (4.15).

- *Funções de influência parciais dos vectores próprios.*

Para a demonstração de (4.16), diferencia-se a equação (4.19), com a aplicação da regra da derivação da função composta, o que conduz a

$$\begin{aligned}
0 &= \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},m}, F)^T \left( \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell m} \lambda_{\ell j}} \boldsymbol{\Sigma}_\ell \right) \boldsymbol{\beta}_j + \\
&+ \boldsymbol{\beta}_m^T \left( \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell m} \lambda_{\ell j}} \boldsymbol{\Sigma}_\ell \right) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) + \\
&+ \boldsymbol{\beta}_m^T \left\{ \sum_{\ell=1}^k \tau_\ell \frac{\partial}{\partial \varepsilon} \left( \frac{\lambda_{\ell m, \varepsilon, i} - \lambda_{\ell j, \varepsilon, i}}{\lambda_{\ell m, \varepsilon, i} \lambda_{\ell j, \varepsilon, i}} \right) \Big|_{\varepsilon=0} \boldsymbol{\Sigma}_\ell + \tau_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \right\} \boldsymbol{\beta}_j, m \neq j.
\end{aligned}$$

Notando que  $\frac{\partial}{\partial \varepsilon} \left( \frac{\lambda_{\ell m, \varepsilon, i} - \lambda_{\ell j, \varepsilon, i}}{\lambda_{\ell m, \varepsilon, i} \lambda_{\ell j, \varepsilon, i}} \right) \Big|_{\varepsilon=0}$  é não nulo quando  $\ell = i$  vem

$$\begin{aligned}
0 &= \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},m}, F)^T \left( \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell m} \lambda_{\ell j}} \boldsymbol{\Sigma}_\ell \right) \boldsymbol{\beta}_j + \\
&+ \boldsymbol{\beta}_m^T \left( \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell m} \lambda_{\ell j}} \boldsymbol{\Sigma}_\ell \right) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) + \\
&+ \tau_i \frac{\partial}{\partial \varepsilon} \left( \frac{\lambda_{im, \varepsilon, i} - \lambda_{ij, \varepsilon, i}}{\lambda_{im, \varepsilon, i} \lambda_{ij, \varepsilon, i}} \right) \Big|_{\varepsilon=0} \boldsymbol{\beta}_m^T \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j + \\
&+ \tau_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \boldsymbol{\beta}_m^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j, \quad \text{para } m \neq j. \quad (4.27)
\end{aligned}$$

Usando de novo  $\boldsymbol{\Sigma}_\ell \boldsymbol{\beta}_j = \lambda_{\ell j} \boldsymbol{\beta}_j$  em (4.23) em conjunto com a condição de ortogonalidade  $\boldsymbol{\beta}_m^T \boldsymbol{\beta}_j = 0$ , para  $m \neq j$ , e a relação (4.19), vem

$$\begin{aligned}
0 &= \left[ \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell m}} \right] [-\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_m] + \\
&+ \left[ \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell j}} \right] \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_m + \\
&+ \tau_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \boldsymbol{\beta}_m^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j = \quad \text{para } m \neq j \\
&= \left[ \sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right] \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_m + \\
&+ \tau_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} \boldsymbol{\beta}_m^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j, \quad \text{para } m \neq j,
\end{aligned}$$

o que implica que

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F)^T \boldsymbol{\beta}_m = \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im} \lambda_{ij}} \frac{\tau_i}{\sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}}} \{ \boldsymbol{\beta}_m^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j \}, m \neq j. \quad (4.28)$$

Estas  $(p - 1)$  equações juntamente com (4.21) podem ser escritas como

$$\boldsymbol{\beta}^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) = \mathbf{u}, \quad (4.29)$$

onde  $\mathbf{u}$  é um vector cuja  $m$ -ésima componente é  $\delta_{mj} c_{mj} \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_m$  e  $c_{mj}$  representa o factor constante dependente dos valores próprios e de  $\tau_i$  de (4.28). Pré-multiplicando (4.29) por  $\boldsymbol{\beta}$  obtém-se

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F) = \sum_{m \neq j} c_{mj} \boldsymbol{\beta}_m \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_m, \quad (4.30)$$

o que escrito na forma vectorial dá (4.16).  $\square$

Como se pode observar, as *funções de influência parciais* de ambos os funcionais dependem da *função de influência* do funcional  $\mathbf{V}_i$ . O seguinte lema caracteriza de uma forma geral a *função de influência* de um funcional *equivariante*,  $\mathbf{V}_i(F_i)$ , da matriz de dispersão.

**Lema 4.1 (caracterização da  $\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i)$ , Croux e Haesbroeck, 2000)** Para qualquer funcional  $\mathbf{V}_i(F_i)$  da matriz de dispersão que goze de *equivariância afim* e possua *função de influência*, existem duas funções  $\alpha_{\mathbf{V}_i}, \gamma_{\mathbf{V}_i} : [0, \infty] \rightarrow \mathbb{R}$  tais que

$$\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) = \alpha_{\mathbf{V}_i} [d_i(\mathbf{x})] \mathbf{x} \mathbf{x}^T - \gamma_{\mathbf{V}_i} [d_i(\mathbf{x})] \boldsymbol{\Sigma}_i, \quad (4.31)$$

com  $d_i^2(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}_i \mathbf{x}$  e  $F_i$  uma distribuição elipticamente simétrica com  $\boldsymbol{\mu}_i = \mathbf{0}_p$ .

**Dem:** Considere-se que  $\mathbf{X}_i \sim F_i$  elipticamente simétrica, com média  $\mathbf{0}_p$  e matriz de dispersão  $\boldsymbol{\Sigma}_i$  e represente-se por  $G$  a distribuição esfericamente simétrica comum a todos os  $\mathbf{Z}_i = \boldsymbol{\Sigma}_i^{-\frac{1}{2}} \mathbf{X}_i$ , para  $i = 1, \dots, k$ . Como  $\mathbf{V}_i$  é um funcional *equivariante para transformações afins*, vem

$$\mathbf{V}_i(\mathbf{X}_i) = \boldsymbol{\Sigma}_i^{\frac{1}{2}} \mathbf{V}_i(\mathbf{Z}_i) \boldsymbol{\Sigma}_i^{\frac{1}{2}},$$

resultando

$$\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) = \boldsymbol{\Sigma}_i^{\frac{1}{2}} \text{IF}(\mathbf{z}_i, \mathbf{V}_i, G) \boldsymbol{\Sigma}_i^{\frac{1}{2}}, \quad (4.32)$$

com  $\mathbf{z}_i = \boldsymbol{\Sigma}_i^{-\frac{1}{2}} \mathbf{x}_i$ .

Note-se que, neste caso a notação para referir os funcionais foi diferente da usada anteriormente. Quando os funcionais são  $p$ -dimensionais é muitas vezes feito este abuso de notação indicando  $\mathbf{V}(\mathbf{X}_i)$  em vez de  $\mathbf{V}(F_i)$ .

Uma vez que  $G$  é uma distribuição esférica e simétrica pelo Lema 1 de Hampel *et al.* (1986), página 276, existem duas funções reais  $\alpha_{\mathbf{V}_i}$  e  $\gamma_{\mathbf{V}_i}$  para as quais

$$\text{IF}(\mathbf{z}, \mathbf{V}_i, G) = \alpha_{\mathbf{V}_i}(\|\mathbf{z}\|) \mathbf{z}\mathbf{z}^T - \gamma_{\mathbf{V}_i}(\|\mathbf{z}\|) \mathbf{I}_p, \quad (4.33)$$

sendo  $\mathbf{I}_p$  a matriz identidade ( $p \times p$ ) e  $\|\cdot\|$  o operador norma Euclidiana. Substituindo (4.33) em (4.32) vem

$$\text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) = \Sigma_i^{\frac{1}{2}} \{ \alpha_{\mathbf{V}_i}(\|\mathbf{z}_i\|) \mathbf{z}_i \mathbf{z}_i^T - \gamma_{\mathbf{V}_i}(\|\mathbf{z}_i\|) \mathbf{I}_p \} \Sigma_i^{\frac{1}{2}}. \quad (4.34)$$

Notando que na expressão anterior  $\|\mathbf{z}_i\| = \left\| \Sigma_i^{-\frac{1}{2}} \mathbf{x} \right\| = \mathbf{x}^T \Sigma_i^{-1} \mathbf{x} = d_i(\mathbf{x})$ , imediatamente resulta (4.31).  $\square$

Em Croux e Haesbroeck (2000) encontra-se indicada a função  $\alpha_{\mathbf{V}_i}$  para os casos em que o funcional  $\mathbf{V}_i(F_i)$  é equivalente ao seguintes estimadores robustos: estimador-M, estimador-S e estimador RMCD. Para o funcional clássico  $\mathbf{S}_i(F_i)$ , i.e. o funcional equivalente à matriz de covariâncias amostral não corrigida, ambas as funções são a identidade.

É interessante constatar que as *funções de influência parciais* para os funcionais dos valores próprios são análogas às obtidas por Croux e Haesbroeck (2000) para o caso de uma população. No que se refere às *funções de influência parciais* dos funcionais vectores próprios, estas têm forma semelhante às deduzidas por Croux e Haesbroeck (2000) para o modelo das *componentes principais*. De facto, estas últimas podem obter-se como caso particular de (4.16) com  $k = 1$

$$\text{IF}(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V}, j}, F) = \sum_{m \neq j} \frac{1}{\lambda_j - \lambda_m} \{ \boldsymbol{\beta}_j^T \text{IF}(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_m \} \boldsymbol{\beta}_m. \quad (4.35)$$

Confrontando (4.16) com (4.35) nota-se que o aumento do número de populações originou a inclusão de pesos, relacionados com os inversos das variâncias dos vectores próprios, os quais dependem dos valores próprios das  $k$  populações. É ainda importante referir que, quando o desvio em relação à esfericidade é nulo para todos os pares de vectores próprios de todos os grupos, as *funções de influência parciais* não estão definidas. Um problema semelhante ocorre com a *função de influência* dos funcionais dos vectores próprios do modelo das *componentes principais*, como

facilmente se observa em (4.35). Esta situação é consequência das observações projectadas, em todos os subespaços gerados por todos os pares de vectores próprios, descreverem esferas perfeitas.

Assim, temos que exigir que em pelo menos uma das populações exista pelo menos um par de valores próprios distintos, para garantir a construção destas funções. No entanto, a exigência (um pouco mais forte) de  $\lambda_{11} > \dots > \lambda_{1p}$  também é necessária no modelo das *componentes principais comuns* para permitir a construção dos vectores comuns sem qualquer problema. Portanto, a restrição referida para construção das *funções de influência parciais* é inerente ao modelo em que trabalhamos.

Da observação das expressões (4.15) e (4.16) é fácil concluir que, com a escolha de estimadores das matrizes de dispersão *B-robustos* os estimadores “plug-in” dos vectores próprios comuns e respectivas variâncias associadas possuem também esta desejada característica. Em contrapartida, verifica-se a ausência desta propriedade quando o funcional  $\mathbf{V}_i$  é equivalente ao estimator clássico  $\mathbf{S}_i$ , a matriz de covariâncias amostral corrigida. Neste caso, e atendendo ao resultado (4.4), as *funções de influência parciais* vêm

$$\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{S}, \ell j}, F) = \delta_{\ell i} \boldsymbol{\beta}_j^T (\mathbf{x} \mathbf{x}^T - \boldsymbol{\Sigma}_i) \boldsymbol{\beta}_j = \delta_{\ell i} \left( (\boldsymbol{\beta}_j^T \mathbf{x})^2 - \lambda_{ij} \right), \quad (4.36)$$

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{S}, j}, F) &= \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im} \lambda_{ij}} \left\{ \sum_{\ell=1}^k \frac{\tau_{\ell} (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right\}^{-1} \times \\ &\times \{ \boldsymbol{\beta}_j^T (\mathbf{x} \mathbf{x}^T - \boldsymbol{\Sigma}_i) \boldsymbol{\beta}_m \} \boldsymbol{\beta}_m = \\ &= \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im} \lambda_{ij}} \left\{ \sum_{\ell=1}^k \frac{\tau_{\ell} (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right\}^{-1} \times \\ &\times (\boldsymbol{\beta}_j^T \mathbf{x} \boldsymbol{\beta}_m^T \mathbf{x}) \boldsymbol{\beta}_m, \end{aligned} \quad (4.37)$$

onde a presença do termo  $\mathbf{x}$ , sem qualquer peso, evidencia a ausência de *B-robustez* do estimator  $\mathbf{S}_i$ . No caso diagonal, isto é quando  $\boldsymbol{\beta} = \mathbf{I}_p$ , a expressão (4.37) escreve-se

$$\begin{aligned} \text{PIF}_i(\mathbf{z}, \boldsymbol{\beta}_{\mathbf{S}, mj}, G) &= \tau_i \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{ij} \lambda_{im}} \left\{ \sum_{\ell=1}^k \frac{\tau_{\ell} (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right\}^{-1} z_m z_j, \quad m \neq j, \\ \text{PIF}_i(\mathbf{z}, \boldsymbol{\beta}_{\mathbf{S}, mj}, G) &= 0, \quad m = j. \end{aligned}$$

No trabalho de Croux e Haesbroeck (2000) encontram-se gráficos das funções (4.36). No que se refere aos vectores próprios, nos gráficos de Croux e Haesbroeck (2000) não intervêm os pesos da equação (4.37). No entanto, a forma das *funções de influência*



*parciais* (4.37) é a mesma das estudadas por Croux e Haesbroeck (2000), e daí que não se tenham incluído gráficos neste trabalho. Além disso, Croux e Haesbroeck (2000) ainda ilustram graficamente o comportamento das *funções de influência* dos funcionais vectores e valores próprios para o caso em que o funcional da matriz de dispersão é o relacionado com o estimador-S.

Como referimos anteriormente, as *funções de influência parciais* podem ser bastante úteis para o cálculo das variâncias assintóticas dos estimadores do modelo em questão. Tendo em conta o resultado (4.6), facilmente obtemos o seguinte corolário do Teorema 4.1, indicando as variâncias assintóticas dos estimadores “plug-in” para o modelo das *componentes principais comuns*.

**Corolário 4.1 (variâncias assintóticas dos estimadores “plug-in” do modelo CPC)**

Seja  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, \dots, k$ , uma amostra aleatória da distribuição  $F_i$ , com parâmetro de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matriz de dispersão  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$ , tal que  $\boldsymbol{\Sigma}_i$  satisfaz o modelo (2.3) com  $\boldsymbol{\beta} = \mathbf{I}_p$ . Assuma-se que  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e  $\sum_{i=1}^k \tau_i = 1$ .

Considerem-se os vectores aleatórios estandardizados  $\boldsymbol{\Lambda}_i^{-\frac{1}{2}} \mathbf{X}_i = \mathbf{Z}_i$  todos com a mesma distribuição esférica  $G$  para todo  $i = 1, \dots, k$ . Assuma-se que,  $\sigma(G_0) = 1$  sendo  $G_0$  a distribuição de  $Z_{1j}$ , com  $j = 1, \dots, p$ , e  $\sigma$  representa o funcional de escala correspondente à raiz quadrada de um elemento da diagonal de  $\mathbf{V}_i(F)$  e  $\sigma_{12}$  o funcional correspondente a um elemento não diagonal de  $\mathbf{V}_i(F)$ . Assuma-se ainda que  $\lambda_{11} > \dots > \lambda_{1p}$  e que existem as variâncias assintóticas de  $\hat{\sigma}$  e  $\hat{\sigma}_{12}$  em  $G_0$ .

Então, as variâncias assintóticas dos estimadores de “plug-in”, solução de (4.10), (4.10) e (4.11) são dadas por

$$\begin{aligned} \text{ASVAR}(\hat{\lambda}_{\ell j}) &= 4\lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(\hat{\sigma}, G_0) , \\ \text{ASVAR}(\hat{\beta}_{jm}) &= \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij} \lambda_{im}} \right]^{-1} \text{ASVAR}(\hat{\sigma}_{12}, G_0 \times G_0) , \quad \text{para } m \neq j , \\ \text{ASVAR}(\hat{\beta}_{jm}) &= 0 , \quad \text{para } m = j , \\ \text{ASCOV}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) &= \sum_{i=1}^k \tau_i \frac{(\lambda_{im} - \lambda_{ij})(\lambda_{ir} - \lambda_{ij})}{\sqrt{\lambda_{im}} \sqrt{\lambda_{ir}}} \times \\ &\quad \times \left[ \sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right]^{-1} \left[ \sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell r} - \lambda_{\ell j})^2}{\lambda_{\ell r} \lambda_{\ell j}} \right]^{-1} E_G[Z_{1j}^2 Z_{1m} Z_{1r}] . \end{aligned}$$

Em particular quando  $G \sim N(\mathbf{0}_p, \mathbf{I}_p)$ , tem-se que

$$\begin{aligned}\text{ASVAR}(\hat{\lambda}_{\ell j}) &= 4\lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(\hat{\sigma}, \Phi) , \\ \text{ASVAR}(\hat{\beta}_{jm}) &= \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij} \lambda_{im}} \right]^{-1} \text{ASVAR}(\hat{\sigma}_{12}, \Phi \times \Phi) , \quad \text{para } m \neq j, \\ \text{ASCOV}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) &= 0 , \quad \text{para } m \neq j \quad m \neq r \quad r \neq j .\end{aligned}$$

**Dem:** A partir do resultado (4.6) e da equação (4.15), obtém-se

$$\begin{aligned}\text{ASVAR}(\hat{\lambda}_{\ell j}) &= \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} \left[ \delta_{\ell i} \boldsymbol{\beta}_j^T \text{IF}(\mathbf{X}, \mathbf{V}_i, F) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T \text{IF}(\mathbf{X}, \mathbf{V}_i, F)^T \boldsymbol{\beta}_j \right] = \\ &= \frac{1}{\tau_\ell} E_{F_\ell} \left[ \boldsymbol{\beta}_j^T \text{IF}(\mathbf{X}, \mathbf{V}_\ell, F) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T \text{IF}(\mathbf{X}, \mathbf{V}_\ell, F)^T \boldsymbol{\beta}_j \right] .\end{aligned}$$

Como  $\frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\lambda_{\ell j}^{\frac{1}{2}}} \sim G_0$  quando  $\mathbf{X} \sim F_i$ , em conjunto com o resultado do Lema 4.1 resulta

$$\begin{aligned}\text{ASVAR}(\hat{\lambda}_{\ell j}) &= \frac{1}{\tau_\ell} E_{F_\ell} \left\{ \left[ \boldsymbol{\beta}_j^T (\alpha_{\mathbf{V}_\ell} \mathbf{X} \mathbf{X}^T - \gamma_{\mathbf{V}_\ell} \boldsymbol{\Sigma}_\ell) \boldsymbol{\beta}_j \right]^2 \right\} = \\ &= \frac{1}{\tau_\ell} E_{F_\ell} \left\{ \left[ \left( \alpha_{\mathbf{V}_\ell} (\boldsymbol{\beta}_j^T \mathbf{X})^2 - \gamma_{\mathbf{V}_\ell} \lambda_{\ell j} \right) \right]^2 \right\} = \\ &= \frac{1}{\tau_\ell} \lambda_{\ell j}^2 E_{F_\ell} \left\{ \left[ \left( \alpha_{\mathbf{V}_\ell} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{X}}{\sqrt{\lambda_{\ell j}}} \right)^2 - \gamma_{\mathbf{V}_\ell} \right) \right]^2 \right\} = \\ &= \frac{1}{\tau_\ell} \lambda_{\ell j}^2 E_{F_\ell} \left[ \text{IF}(\mathbf{Z}_i, \sigma^2, G_0)^2 \right] = \\ &= 4\lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(\hat{\sigma}, G_0) .\end{aligned}$$

Para o estimador dos vectores próprios, a variância assímtótica é dada por

$$\text{ASVAR}(\hat{\beta}_{mj}) = \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} \left[ \text{PIF}_i(\mathbf{X}, \beta_{\mathbf{V}, mj}, F)^2 \right] .$$

Para  $(m = j)$  o resultado é óbvio porque  $\text{PIF}_i(\mathbf{x}, \beta_{\mathbf{V}, mj}, F) = 0$ .

Para  $(m \neq j)$  do Lema 4.1 e de (4.16) para o caso diagonal tem-se

$$\text{PIF}_i(\mathbf{x}, \beta_{\mathbf{V}, mj}, F) = \tau_i \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im} \lambda_{ij}} \left\{ \sum_{\ell=1}^k \frac{\tau_\ell (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m} \lambda_{\ell j}} \right\}^{-1} \alpha_{\mathbf{V}_i} [d_i(\mathbf{x})] \boldsymbol{\beta}_j^T \mathbf{x} \boldsymbol{\beta}_m^T \mathbf{x}^T . \quad (4.38)$$

Logo

$$\begin{aligned} \text{ASVAR}(\widehat{\beta}_{mj}) &= \sum_{i=1}^k \tau_i \left( \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im}\lambda_{ij}} \right)^2 \left\{ \sum_{\ell=1}^k \frac{\tau_\ell (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m}\lambda_{\ell j}} \right\}^{-2} \times \\ &\times E_{F_i} \left\{ \left[ \alpha_{\mathbf{V}_i} [d_i(\mathbf{X})] \beta_j^T \mathbf{X} \beta_m^T \mathbf{X}^T \right]^2 \right\}, \quad \text{para } m \neq j. \end{aligned}$$

Assim, tem-se que

$$\begin{aligned} \text{ASVAR}(\widehat{\beta}_{mj}) &= \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{im}\lambda_{ij}} \left\{ \sum_{\ell=1}^k \frac{\tau_\ell (\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m}\lambda_{\ell j}} \right\}^{-2} \times \\ &\times E_{F_i} \left\{ \left[ \alpha_{\mathbf{V}_i} [d_i(\mathbf{X})] \frac{\beta_j^T \mathbf{X}}{\sqrt{\lambda_{ij}}} \frac{\beta_m^T \mathbf{X}^T}{\sqrt{\lambda_{\ell m}}} \right]^2 \right\}, \quad \text{para } m \neq j, \end{aligned}$$

isto é

$$\text{ASVAR}(\widehat{\beta}_{mj}) = \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij}\lambda_{im}} \right]^{-1} E_G \{ \text{IF}(\mathbf{Z}, \sigma_{12}, G_0) \}^2, \quad \text{para } m \neq j,$$

e portanto

$$\text{ASVAR}(\widehat{\beta}_{mj}) = \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij}\lambda_{im}} \right]^{-1} \text{ASVAR}(\hat{\sigma}_{12}, G_0), \quad \text{para } m \neq j,$$

como queríamos demonstrar.

Para a covariância assintótica, sabemos que

$$\text{ASCOV}(\widehat{\beta}_{jm}, \widehat{\beta}_{jr}) = \sum_{m \neq j \neq r} \frac{1}{\tau_i} E_{F_i} [\text{PIF}_i(\mathbf{X}, \beta_{\mathbf{V}, jm}, F) \text{PIF}_i(\mathbf{X}, \beta_{\mathbf{V}, jr}, F)].$$

Utilizando a expressão (4.16) e o facto de  $\frac{\mathbf{X}^T \beta_j}{\lambda_{\ell j}^{\frac{1}{2}}} \sim G_0$  quando  $\mathbf{X} \sim F_i$ , resulta

$$\begin{aligned} \text{ASCOV}(\widehat{\beta}_{jm}, \widehat{\beta}_{jr}) &= \sum_{i=1}^k \tau_i \frac{(\lambda_{im} - \lambda_{ij})(\lambda_{ir} - \lambda_{ij})}{\sqrt{\lambda_{im}} \sqrt{\lambda_{ir}}} \times \\ &\times \left[ \sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m}\lambda_{\ell j}} \right]^{-1} \left[ \sum_{\ell=1}^k \tau_\ell \frac{(\lambda - \lambda)^2}{\lambda\lambda} \right]^{-1} \times \\ &\times E_G [Z_{1j}^2 Z_{1m} Z_{1r}], \end{aligned} \quad (4.39)$$

quando  $m \neq j$ ,  $m \neq r$  e  $r \neq j$ .

No caso particular de  $G \sim N(\mathbf{0}_p, \mathbf{I}_p)$ , como  $Z_{1j}$ ,  $Z_{1m}$  e  $Z_{1r}$  são variáveis aleatórias independentes, isto conduz à anulação do segundo membro da equação (4.39).  $\square$

Os resultados apresentados no Corolário 4.1 também foram derivados por Boente e Orellana (2001) para o caso particular das variáveis terem distribuição normal  $p$ -variada. No entanto, a abordagem utilizada por Boente e Orellana (2001) para obter estes resultados foi diferente da aqui considerada.

Como caso especial do Corolário 4.1, vamos considerar o funcional  $\mathbf{V}_i(F_i) = \mathbf{S}_i(F_i)$  para populações normalmente distribuídas. Como se viu no Capítulo 3 a *função de influência* do funcional variância é

$$\text{IF}(x, \sigma^2(F), F) = x^2 - \sigma^2, \quad (4.40)$$

quando  $E(X) = 0$ .

Assim, a variância assintótica dos estimadores dos valores próprios fica

$$\text{ASVAR}(\hat{\lambda}_{ij}) = \frac{4\lambda_{ij}^2}{\tau_i} E_G [\text{IF}(Z, \sigma, G_0)^2] = \frac{4\lambda_{ij}^2}{\tau_i} E_G \left( \frac{Z_{11}^2 - 1}{2} \right)^2. \quad (4.41)$$

Note-se que na equação (4.41) se aplicou a regra de derivação da função composta para obter

$$\text{IF}(x, \sigma(F), F) = \frac{\text{IF}(x, \sigma^2(F), F)}{2\sigma(G_0)}.$$

O facto de  $\sigma(G_0) = 1$ , conduz a

$$\text{ASVAR}(\hat{\lambda}_{ij}) = \frac{\lambda_{ij}^2}{\tau_i} E_G [Z_{11}^4 - 1].$$

Tendo em conta a expressão do  $r$ -ésimo momento ordinário da distribuição normal univariada

$$E(X^r) = (r-1)(r-3)\dots 1 \times \sigma^r, \quad \text{para } r \text{ par},$$

vem que

$$\text{ASVAR}(\hat{\lambda}_{ij}) = \frac{2\lambda_{ij}^2}{\tau_i},$$

sendo esta a variância assintótica dos estimadores de máxima verosimilhança obtida por Flury (1986, 1988, Capítulo 4).

No que se refere ao estimador dos vectores próprios, temos que

$$\text{ASVAR}(\hat{\beta}_{jm}) = \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij} \lambda_{im}} \right]^{-1} E_G [\text{IF}(\mathbf{Z}, \sigma_{12}, G_0)]^2, \quad m \neq j, \quad (4.42)$$

com  $\sigma_{12}$  o funcional correspondente a um elemento não diagonal de  $\mathbf{V}_i(F)$ .

Analogamente a (4.40), temos a seguinte *função de influência* para o funcional  $\sigma_{12}$ , equivalente ao elemento  $S_{12}$  da matriz de covariâncias amostral não corrigida  $\mathbf{S}_i$

$$\text{IF}(\mathbf{x}, \sigma_{12}(F), F) = x_1 x_2 - \sigma_{12}, \quad (4.43)$$

onde  $\sigma_{12}$  representa o elemento correspondente da matriz  $\Sigma_i$ . A aplicação deste resultado em (4.42) conduz a

$$\text{ASVAR}(\hat{\beta}_{jm}) = \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij} \lambda_{im}} \right]^{-1} E_G[Z_{11}^2 Z_{12}^2], \quad \text{para } m \neq j.$$

Como  $\sigma(G_0) = 1$ , vem

$$\text{ASVAR}(\hat{\beta}_{jm}) = \left[ \sum_{i=1}^k \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij} \lambda_{im}} \right]^{-1},$$

o que coincide com a variância assintótica do estimador de máxima verosimilhança dos vectores próprios comuns (Flury, 1986, 1988, Capítulo 4), como não podia deixar de ser.

Com as variâncias assintóticas deduzidas anteriormente, e seguindo o trabalho de Croux e Haesbroeck (2000), é possível definir as eficiências assintóticas relativas dos estimadores sob o modelo normal como

$$\text{Eff}(\hat{\lambda}_{ij}) = \frac{\text{ASVAR}(\hat{\lambda}_{ij})|_{\mathbf{V}=\mathbf{S}}}{\text{ASVAR}(\hat{\lambda}_{ij})} = \frac{1}{2\text{ASVAR}(\hat{\sigma}, \Phi)} = \frac{2}{\text{ASVAR}(\hat{\sigma}^2, \Phi)}, \quad (4.44)$$

e

$$\text{Eff}(\hat{\beta}_{jm}) = \frac{\text{ASVAR}(\hat{\beta}_{jm})|_{\mathbf{V}=\mathbf{S}}}{\text{ASVAR}(\hat{\beta}_{jm})} = \frac{1}{\text{ASVAR}(\hat{\sigma}_{12}, \Phi)}. \quad (4.45)$$

Como se verifica em (4.45) as eficiências assintóticas dos estimadores dos vectores próprios comuns dependem da eficiência dos elementos de  $\mathbf{V}$  não diagonais. Por sua vez, para os valores próprios a eficiência apresentada em (4.44) relaciona-se com a eficiência dos elementos diagonais de  $\mathbf{V}$ .

### 4.3.2 Funções de influência parciais e variâncias assintóticas dos estimadores “projection-pursuit”

O método de estimação “projection-pursuit” (PP) pode ser aplicado em alternativa à metodologia “plug-in” para a estimação dos parâmetros do modelo das *componentes principais comuns*.

A técnica PP foi originalmente proposta por Kruskal (1969, 1972). No entanto, foram os autores Friedman e Tukey (1974) que implementaram com sucesso esta técnica. Aliás, foi nesse mesmo trabalho que pela primeira vez se utilizou a designação de “projection–pursuit”. Esta técnica exploratória quando aplicada a uma nuvem de pontos multidimensionais procura encontrar projecções, num espaço de baixa dimensão (unidimensional ou bidimensional), que revelem características importantes dos dados. Como Huber (1985) refere,

*“PP emerges as the most powerful method yet invented to lift one–dimensional statistical techniques to higher dimensions”.*

Para expormos o método PP, é necessário introduzir a noção de *projecção linear*.

**Definição 4.4 (projecção linear)** Uma *projecção linear* de  $\mathbb{R}^p$  em  $\mathbb{R}^s$  é qualquer aplicação linear do tipo

$$\mathbf{Z} = \mathbf{B}\mathbf{X}, \quad \mathbf{X} \in \mathbb{R}^p, \quad \mathbf{Z} \in \mathbb{R}^s,$$

onde  $\mathbf{B}$  é uma matriz ( $s \times p$ ) com característica  $s$ .

Quando os vectores que constituem as linhas da matriz  $\mathbf{B}$  são ortonormados, a projecção é designada por *projecção ortogonal*.

Sendo  $\mathbf{X}$  uma variável aleatória  $p$ –dimensional com função de distribuição  $F$ , então a variável aleatória projectada,  $s$ –dimensional,  $\mathbf{Z} = \mathbf{B}\mathbf{X}$  tem distribuição referida por  $F[\mathbf{B}]$ . Por definição o método PP procura a projecção  $\mathbf{B}$  maximizando, ou minimizando, uma certa função objectivo ou *índice de projecção*  $\varsigma(F[\mathbf{B}])$ .

No caso particular de  $s = 1$ , quando  $\mathbf{B}$  se reduz a um vector e que referimos por  $\mathbf{b}^T$ , a função de distribuição da variável aleatória univariada  $Z$  é  $F[\mathbf{b}]$ . É muito frequente o emprego da técnica PP neste caso particular, utilizando vectores normados. Isto porque, no caso unidimensional é possível examinar de uma forma mais exaustiva qual a projecção que revela aspectos interessantes. Em altas dimensões este processo é quase impraticável. Além disso, como Huber (1985) refere, uma função de distribuição  $p$ –dimensional fica completamente caracterizada pelo conjunto de todas as projecções unidimensionais  $F[\mathbf{b}]$ . Isto é consequência de qualquer função de distribuição  $F$  (neste caso multivariada) ser unicamente determinada pela sua função característica  $\psi$  e por a função característica  $\psi_{\mathbf{b}}$  de uma projecção unidimensional  $F[\mathbf{b}]$ , na direcção de  $\mathbf{b}$ , igualar o corte de  $\psi$  nessa direcção,

$$\psi_{\mathbf{b}}(t) = E\left(e^{it\mathbf{b}^T\mathbf{X}}\right) = \psi(t\mathbf{b}).$$

Pela forma como o método PP é definido, é fácil apercebermo-nos que alguns métodos clássicos da análise multivariada são casos especiais da técnica PP. Entre eles podemos citar a análise em *componentes principais* e a *análise discriminante*.

Por outro lado, quando o método de estimação PP é baseado num *índice de projecção* robusto, consegue-se obter estimadores que aliam boas propriedades de robustez e eficiência. Como Donoho *et al.* (1985) salientam o *alto ponto de ruptura* está dependente do comportamento do estimador em projecções especiais. Como a técnica PP é utilizada para pesquisar as projecções apropriadas, e actuar (proceder com o cálculo do *índice de projecção*) quando as encontra, é natural que preserve esta propriedade. Além disso, com o método PP é possível criar estimadores *equivariantes*. Como exemplo podemos referir o estimador de Stahel–Donoho (SDE), que se baseia em projecções e combina as propriedades *alto ponto de ruptura* e *equivariância para transformações afins*, como vimos no Capítulo 3.

Inicialmente, a estimação via PP apresentava o inconveniente de consumir muito tempo computacional. É fácil apercebermo-nos que a pesquisa das direcções que revelam características interessantes dos dados poderá ser bastante dispendiosa. Actualmente, com o avanço da tecnologia informática este problema perdeu parte da sua importância e a técnica PP é cada vez mais utilizada.

No que se refere ao modelo das *componentes principais*, a proposta da estimação das direcções principais e respectivas variâncias via PP foi da autoria de Li e Chen (1985). Estes autores construíram estimadores PP através da maximização (e minimização) iterativa de um estimador de escala robusto. Verificaram ainda que estes estimadores são *equivariantes para transformações ortogonais*, consistentes, gozam de *robustez qualitativa* e mantêm o *ponto de ruptura* dos estimadores de escala. Croux e Ruiz–Gazen (1996) desenvolveram um algoritmo rápido para o cálculo destes estimadores e posteriormente Croux e Ruiz–Gazen (2000) estudaram o *ponto de ruptura* das matrizes de dispersão relacionadas com os estimadores obtidos com este algoritmo. Além disso, Croux e Ruiz–Gazen (2000) ainda obtiveram as *funções de influência* dos funcionais “projection–pursuit”. Dada a relevância deste tema, vários outros autores aplicaram e estudaram esta metodologia, por exemplo, Patak (1991), Xie *et al.* (1993), Berrendero (1996) e Hubert *et al.* (2002). Muito recentemente, Cui *et al.* (2002) deduziram a distribuição assintótica dos estimadores PP das *componentes principais*.

Como referimos na Secção 2, Boente e Orellana (2001) aplicaram a metodologia PP na estimação dos vectores próprios comuns, e respectivas variâncias associadas para cada grupo, a dados cuja estrutura de covariâncias é bem traduzida pelo modelo

CPC. Para patentear a metodologia utilizada, considera-se  $\mathbf{X}_{i\ell}$ ,  $i = 1, \dots, k$  e  $\ell = 1, \dots, n_i$ , vectores aleatórios independentes de  $k$  amostras independentes de  $\mathbb{R}^p$  com parâmetro de localização  $\boldsymbol{\mu}_i$  e matriz de dispersão  $\boldsymbol{\Sigma}_i$ , com  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i})$ . Novamente vamos assumir, para simplificar a exposição e sem perda de generalidade, que  $\boldsymbol{\mu}_i = \mathbf{0}_p$ .

Seguindo as ideias de Li e Chen (1985) para o caso de uma população, os estimadores das direcções comuns  $\boldsymbol{\beta}$  são solução de

$$\hat{\boldsymbol{\beta}}_1 = \arg \max_{\|\mathbf{b}\|=1} \sum_{i=1}^k \tau_i s^2(\mathbf{X}_i^T \mathbf{b}), \quad \hat{\boldsymbol{\beta}}_j = \arg \max_{\mathbf{b} \in \mathcal{B}_j} \sum_{i=1}^k \tau_i s^2(\mathbf{X}_i^T \mathbf{b}), \quad j = 2, \dots, p, \quad (4.46)$$

onde  $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}^T \hat{\boldsymbol{\beta}}_m = 0, \text{ para } m = 1, \dots, j-1\}$  e  $s$  representa um estimador de escala univariado. Os pesos  $\tau_i$  foram incluídos de forma a adaptar o procedimento às diferentes dimensões amostrais.

Os estimadores dos valores próprios da  $i$ -ésima população são então calculados com  $\hat{\lambda}_{ij} = s^2(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_j)$ , para  $j = 1, \dots, p$ .

Este procedimento baseia-se no facto de para o modelo das CPC a decomposição comum dada em (2.3) implicar que para qualquer  $\mathbf{b} \in \mathbb{R}^p$ , e  $i = 1, \dots, k$ , a  $\text{VAR}(\mathbf{b}^T \mathbf{X}_{i1}) = \mathbf{b}^T \boldsymbol{\beta} \boldsymbol{\Lambda}_i \boldsymbol{\beta}^T \mathbf{b}$ . Assim, o primeiro (ou o último) eixo principal  $\boldsymbol{\beta}_1$  pode ser obtido com a maximização (ou a minimização) de  $\sum_{i=1}^k \tau_i \text{VAR}(\mathbf{b}^T \mathbf{X}_{i1})$  com  $\mathbf{b} \in \mathbb{R}^p$  e  $\|\mathbf{b}\| = 1$ . Para permitir este procedimento ainda é necessário exigir que a matriz  $\boldsymbol{\Sigma}_{\text{id}} = \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i$  não tenha valores próprios múltiplos. O segundo eixo principal é definido considerando direcções ortogonais a  $\boldsymbol{\beta}_1$  e assim sucessivamente. Os valores próprios da  $i$ -ésima população são claramente  $\text{VAR}(\mathbf{X}_i^T \boldsymbol{\beta}_j) = \boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j = \lambda_{ij}$ . Finalmente, ordenam-se os vectores próprios de acordo com a ordenação decrescente dos valores próprios da primeira população. Isto permite definir os estimadores “projection-pursuit” para o modelo CPC como em (4.46).

Como salientámos no capítulo anterior, a *equivariância afim* é uma propriedade desejável. Quanto aos estimadores PP das CPC, estes possuem *equivariância ortogonal*, isto é, são *equivariantes* quando é aplicada a mesma matriz de rotação ortogonal a todas as populações. A exigência da mesma rotação preserva a estrutura de covariâncias do modelo das CPC como referem Boente e Orellana (2001).

Para o cálculo destes estimadores Boente e Orellana (2001) generalizam o algoritmo de Croux e Ruiz-Gazen (1996), que calcula a proposta de Li e Chen (1985) para problemas com  $k$  populações. No entanto, as autoras Boente e Orellana (2001) não



deduziram as *funções de influência parciais* associadas aos funcionais equivalentes aos estimadores de interesse. Só com a construção destas funções é que o trabalho de Croux e Ruiz-Gazen (2000) será generalizado para situações onde intervêm vários grupos. Neste sentido, nesta subsecção vamos construir as *funções de influência parciais* dos funcionais vectores próprios comuns e valores próprios, como a adicional dedução das variâncias dos estimadores equivalentes. Para isto, considera-se que  $\mathbf{X}_i$  são vectores aleatórios independentes, tais que  $\mathbf{X}_i \sim F_i$ , onde  $F_i$  tem parâmetro de localização  $\mathbf{0}_p$  e matriz de dispersão  $\Sigma_i = \mathbf{C}_i \mathbf{C}_i^T$  que satisfaz (2.3), e representa-se por  $F_i[\mathbf{b}]$  a distribuição de  $\mathbf{b}^T \mathbf{X}_i$ .

Seja  $\varsigma(\mathbf{b}) = \varsigma(F[\mathbf{b}]) = \sum_{i=1}^k \tau_i \sigma^2(F_i[\mathbf{b}])$  onde  $\sigma(\cdot)$  é um funcional de escala univariado, o qual se supõe *equivariante* em escala.

Considere-se a distribuição  $F$  definida pela medida produto,  $F = F_1 \times \cdots \times F_k$ . Os funcionais PP das direcções comuns  $\boldsymbol{\beta}_\sigma(F) = (\boldsymbol{\beta}_{\sigma,1}(F), \dots, \boldsymbol{\beta}_{\sigma,p}(F))$  são solução de

$$\begin{aligned} \boldsymbol{\beta}_{\sigma,1}(F) &= \arg \max_{\|\mathbf{b}\|=1} \varsigma(\mathbf{b}) , \\ \boldsymbol{\beta}_{\sigma,j}(F) &= \arg \max_{\mathbf{b} \in \mathcal{B}_j} \varsigma(\mathbf{b}) , \quad j = 2, \dots, p , \end{aligned}$$

onde  $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}^T \boldsymbol{\beta}_m(F) = 0, \text{ para } m = 1, \dots, j-1\}$ .

Por sua vez, os funcionais relacionados com os estimadores dos valores próprios e da matriz de dispersão são definidos por

$$\lambda_{\sigma,ij}(F) = \sigma^2(F_i[\boldsymbol{\beta}_{\sigma,j}(F)]) , \quad (4.47)$$

$$\mathbf{V}_{\sigma,i}(F) = \sum_{j=1}^p \lambda_{\sigma,ij}(F) \boldsymbol{\beta}_{\sigma,j}(F) \boldsymbol{\beta}_{\sigma,j}(F)^T . \quad (4.48)$$

Embora os estimadores definidos com a maximização e a minimização sejam diferentes os funcionais equivalentes tomam os mesmos valores para distribuições teóricas pelo que as suas *funções de influência parciais* são as mesmas.

Quando a matriz  $\Sigma_{\text{id}} = \sum_{i=1}^k \tau_i \Sigma_i$  não tem valores próprios múltiplos, Boente e Orellana (2001) mostraram que os funcionais  $\boldsymbol{\beta}_\sigma(F)$  e  $\lambda_{\sigma,ij}(F)$  são Fisher-consistentes para qualquer distribuição  $F$ , caso  $\mathbf{Z}_i = \mathbf{C}_i^{-1} \mathbf{X}_i$  tenha a mesma distribuição esférica  $G$ ,  $i = 1, \dots, k$ . Para isto se verificar, ainda é necessário admitir que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição de  $Z_{11}$ . Portanto, caso estas condições sejam satisfeitas, tem-se que  $\boldsymbol{\beta}_\sigma(F) = \boldsymbol{\beta}$  e  $\lambda_{\sigma,ij}(F) = \lambda_{ij}$ , como é desejável.

No teorema seguinte apresentamos as expressões das *funções de influência parciais* dos funcionais PP.

**Teorema 4.2 (funções de influência parciais dos funcionais equivalentes aos estimadores PP do modelo CPC)**

Sejam  $\mathbf{X}_i$  vectores aleatórios independentes com distribuição elipticamente simétrica  $F_i$ , com parâmetro de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matriz de dispersão  $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^T$ , que satisfaz o modelo (2.3). Considerem-se ainda os vectores aleatórios  $\mathbf{Z}_i = \mathbf{C}_i^{-1} \mathbf{X}_i$  com distribuição esférica  $G$  para todo  $i = 1, \dots, k$ . Assuma-se também que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição de  $Z_{11}$ , e que  $\boldsymbol{\Sigma}_{\text{id}} = \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i = \boldsymbol{\beta} \text{diag}(\nu_1, \dots, \nu_p) \boldsymbol{\beta}^T$  com  $\nu_1 > \nu_2 > \dots > \nu_p$ .

Então, se a função  $(\varepsilon, y) \rightarrow \sigma((1 - \varepsilon)G_0 + \varepsilon\Delta_y)$  é contínua e duas vezes diferenciável em  $(0, y)$ , tem-se que para qualquer  $\mathbf{x}$

$$\text{PIF}_i(\mathbf{x}, \lambda_{\sigma, \ell j}, F) = \delta_{\ell i} 2 \lambda_{ij} \text{IF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right), \quad (4.49)$$

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma, j}, F) &= \tau_i \sqrt{\lambda_{ij}} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \sum_{s=j+1}^p \frac{1}{\nu_j - \nu_s} \boldsymbol{\beta}_s (\mathbf{x}^T \boldsymbol{\beta}_s) + \\ &+ \tau_i \sum_{s=1}^{j-1} \frac{1}{\nu_j - \nu_s} \boldsymbol{\beta}_s \sqrt{\lambda_{is}} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j), \end{aligned} \quad (4.50)$$

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma, \ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\nu_j - \nu_s} \tau_i \sqrt{\lambda_{is}} \times \\ &\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) (\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T), \end{aligned} \quad (4.51)$$

onde  $\text{DIF}(y, \sigma, G_0)$  representa a derivada de  $\text{IF}(y, \sigma, G_0)$  em ordem a  $y$ .

**Dem:** A técnica usada nesta demonstração é a mesma da utilizada por Croux e Ruiz-Gazen (2000) na demonstração do Teorema 1.

Represente-se por  $F$  a distribuição definida pela medida produto  $F = F_1 \times \dots \times F_k$ , e assumam-se que  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e  $\sum_{i=1}^k \tau_i = 1$ .

Seja  $F_{i, \mathbf{x}, \varepsilon} = (1 - \varepsilon)F_i + \varepsilon\Delta_{\mathbf{x}}$  com  $F_{\mathbf{x}, \varepsilon, i} = F_1 \times \dots \times F_{i-1} \times F_{i, \mathbf{x}, \varepsilon} \times F_{i+1} \times \dots \times F_k$  a medida produto com contaminação na  $i$ -ésima população. Considerem-se os seguintes funcionais sob contaminação,  $\boldsymbol{\beta}_{j, \varepsilon, i} = \boldsymbol{\beta}_{\sigma, j}(F_{\mathbf{x}, \varepsilon, i})$ ,  $\lambda_{\ell j, \varepsilon, i} = \sigma^2(F_{\ell}[\boldsymbol{\beta}_{j, \varepsilon, i}])$ , para  $\ell \neq i$  e  $\lambda_{ij, \varepsilon, i} = \sigma^2(F_{i, \mathbf{x}, \varepsilon}[\boldsymbol{\beta}_{j, \varepsilon, i}])$ . Seja ainda  $\mathbf{V}_{\ell, \varepsilon, i} = \mathbf{V}_{\sigma, \ell}(F_{\mathbf{x}, \varepsilon, i}) = \sum_{j=1}^p \lambda_{\ell j, \varepsilon, i} \boldsymbol{\beta}_{j, \varepsilon, i} \boldsymbol{\beta}_{j, \varepsilon, i}^T$ .

- *Funções de influência parciais dos vectores próprios.*

Para  $j$  fixo pretende-se obter as *funções de influência parciais* do funcional  $\beta_{j,\varepsilon,i}$ , que são dadas por  $\frac{\partial}{\partial \varepsilon} \beta_{j,\varepsilon,i} \big|_{\varepsilon=0}$ .

O funcional  $\beta_{j,\varepsilon,i}$  é tal que maximiza  $\varsigma(F_{\mathbf{x},\varepsilon,i}[\mathbf{b}])$ , sujeito à restrição de ortogonalidade, i.e.,  $\beta_{j,\varepsilon,i}^T \beta_{j,\varepsilon,i} = 1$  e  $\beta_{s,\varepsilon,i}^T \beta_{j,\varepsilon,i} = 0$  para  $s = 1, \dots, j-1$  e  $j = 1, \dots, p$ . Assim, resulta que  $\beta_{j,\varepsilon,i}$  maximiza a função Lagrangeana

$$L(\mathbf{b}, \gamma, \alpha) = \tau_i \sigma^2(F_{i,\mathbf{x},\varepsilon}[\mathbf{b}]) + \sum_{i_0 \neq i} \tau_{i_0} \sigma^2(F_{i_0}[\mathbf{b}]) - \gamma(\mathbf{b}^T \mathbf{b} - 1) - \sum_{s=1}^{j-1} \alpha_s \mathbf{b}^T \beta_{s,\varepsilon,i}, \quad (4.52)$$

e assim, deve satisfazer a equação

$$\mathbf{0}_p = \frac{\partial}{\partial \mathbf{b}} L(\mathbf{b}, \gamma, \alpha) \big|_{\mathbf{b}=\beta_{j,\varepsilon,i}} = \psi(\varepsilon) - 2\gamma \beta_{j,\varepsilon,i} - \sum_{s=1}^{j-1} \alpha_s \beta_{s,\varepsilon,i}, \quad (4.53)$$

com

$$\begin{aligned} \psi(\varepsilon) &= \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{\mathbf{x},\varepsilon,i}[\mathbf{b}]) \big|_{\mathbf{b}=\beta_{j,\varepsilon,i}} = \\ &= \tau_i \frac{\partial}{\partial \mathbf{b}} \sigma^2(F_{i,\mathbf{x},\varepsilon}[\mathbf{b}]) \big|_{\mathbf{b}=\beta_{j,\varepsilon,i}} + \sum_{i_0 \neq i} \tau_{i_0} \frac{\partial}{\partial \mathbf{b}} \sigma^2(F_{i_0}[\mathbf{b}]) \big|_{\mathbf{b}=\beta_{j,\varepsilon,i_0}}. \end{aligned} \quad (4.54)$$

Como  $\beta_{j,\varepsilon,i}^T \beta_{j,\varepsilon,i} = 1$  e  $\beta_{s,\varepsilon,i}^T \beta_{j,\varepsilon,i} = 0$  para  $s = 1, \dots, j-1$  tem-se que

$$\begin{aligned} \psi(\varepsilon)^T \beta_{j,\varepsilon,i} &= 2\gamma, \\ \psi(\varepsilon)^T \beta_{s,\varepsilon,i} &= \alpha_s, \quad s = 1, \dots, j-1. \end{aligned} \quad (4.55)$$

Consequentemente, (4.53) pode ser escrito como

$$\psi(\varepsilon) = \sum_{s=1}^j (\psi(\varepsilon)^T \beta_{j,\varepsilon,i}) \beta_{j,\varepsilon,i}. \quad (4.56)$$

Derivando (4.56) em ordem a  $\varepsilon$ , vem

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon) \big|_{\varepsilon=0} &= \sum_{s=1}^j [(\psi(0)^T \text{PIF}_i(\mathbf{x}, \beta_{\sigma,s}, F)) \beta_s] + \\ &+ \sum_{s=1}^j \left[ \left( \beta_{\sigma,s}^T \frac{\partial}{\partial \varepsilon} \psi(\varepsilon) \big|_{\varepsilon=0} \right) \beta_s + (\psi(0)^T \beta_s) \text{PIF}_i(\mathbf{x}, \beta_{\sigma,s}, F) \right]. \end{aligned} \quad (4.57)$$

Como a distribuição  $F_i$  é elipticamente simétrica e  $\sigma(G_0) = 1$ , usando o facto de  $\sigma^2(F_i[\mathbf{b}]) = \mathbf{b}^T \Sigma_i \mathbf{b}$ , obtém-se

$$\begin{aligned} \psi(0) &= \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{\mathbf{x},0,i}[\mathbf{b}]) \big|_{\mathbf{b}=\beta_j} = 2 \sum_{i_0=1}^k \tau_{i_0} \Sigma_{i_0} \beta_j = \\ &= 2 \Sigma_{\text{id}} \beta_j = 2\nu_j \beta_j, \end{aligned}$$

o que implica  $\psi(0)^T \boldsymbol{\beta}_s = 0$ , para  $s = 1, \dots, j-1$ .

Representando por  $\mathbf{P}_{j+1} = \mathbf{I}_p - \sum_{s=1}^j \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T$ , então (4.57) pode ser escrita como

$$\mathbf{P}_{j+1} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} = 2\nu_j \sum_{s=1}^j \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s + 2\nu_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F). \quad (4.58)$$

Note-se que o resultado (4.58) é consequência de  $\mathbf{P}_{j+1} \boldsymbol{\beta}_j = \mathbf{0}_p$  e de

$$\mathbf{P}_{j+1} \left( \boldsymbol{\beta}_s^T \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} \right) \boldsymbol{\beta}_s = \mathbf{0}_p.$$

Por outro lado, de (4.54) e uma vez que  $\varsigma(F[\mathbf{b}]) = \mathbf{b}^T \boldsymbol{\Sigma}_{\text{id}} \mathbf{b}$ , pela regra da derivação da função composta vem que

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= \frac{\partial}{\partial \varepsilon} \left( \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{\mathbf{x}, \varepsilon, i}(\mathbf{b})) \right) \Big|_{\mathbf{b}=\boldsymbol{\beta}_{j,i, \varepsilon=0}} + \\ &+ \frac{\partial}{\partial \mathbf{b}} \frac{\partial}{\partial \mathbf{b}^T} \varsigma(F_i(\mathbf{b})|_{\mathbf{b}=\boldsymbol{\beta}_j}) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= 2 \boldsymbol{\Sigma}_{\text{id}} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + \tau_i \frac{\partial}{\partial \mathbf{b}} \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}]) \Big|_{\mathbf{b}=\boldsymbol{\beta}_j}. \end{aligned} \quad (4.59)$$

Usando a equivariância do funcional de escala, vem

$$\begin{aligned} \frac{\partial}{\partial \mathbf{b}} \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}]) \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} &= \frac{\partial}{\partial \mathbf{b}} (\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}) \text{IF} \left( \frac{\mathbf{b}^T \mathbf{x}}{\sqrt{\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}}}, \sigma^2, G_0 \right) \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} = \\ &= 2\lambda_{ij} \boldsymbol{\beta}_j \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) + \\ &+ \lambda_{ij} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \left[ \frac{\mathbf{x}}{\sqrt{\lambda_{ij}}} - \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}} \boldsymbol{\beta}_j \right]. \end{aligned} \quad (4.60)$$

De (4.59) e (4.60) obtém-se

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= 2 \boldsymbol{\Sigma}_{\text{id}} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + 2\tau_i \lambda_{ij} \boldsymbol{\beta}_j \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) + \\ &+ \tau_i \lambda_{ij} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \left[ \frac{\mathbf{x}}{\sqrt{\lambda_{ij}}} - \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}} \boldsymbol{\beta}_j \right]. \end{aligned} \quad (4.61)$$

Portanto de (4.61), (4.58) pode ser escrita como

$$\begin{aligned} &2\nu_j \sum_{s=1}^j \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s + 2\nu_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= 2 \mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + \tau_i \sqrt{\lambda_{ij}} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \mathbf{P}_{j+1} \mathbf{x}. \end{aligned} \quad (4.62)$$

Notando que  $\boldsymbol{\beta}_{j,\varepsilon}^T \boldsymbol{\beta}_{j,\varepsilon} = 1$ , resulta a ortogonalidade entre o funcional vector próprio e as suas *funções de influência parciais*, i.e.,  $\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T \boldsymbol{\beta}_j = 0$ . Com este facto, (4.62) é equivalente a

$$2(\mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} - \nu_j \mathbf{I}_p) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = 2\nu_j \sum_{s=1}^{j-1} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s - \tau_i \sqrt{\lambda_{ij}} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \mathbf{P}_{j+1} \mathbf{x}. \quad (4.63)$$

Por outro lado, como

$$\begin{aligned} \mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} &= \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i - \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i \sum_{s=1}^j \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\ &= \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i - \sum_{i=1}^k \sum_{s=1}^j \tau_i \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\ &= \sum_{i=1}^k \sum_{s=1}^p \tau_i \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T - \sum_{i=1}^k \sum_{s=1}^j \tau_i \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\ &= \sum_{i=1}^k \sum_{s=j+1}^p \tau_i \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\ &= \sum_{s=j+1}^p \left( \sum_{i=1}^k \tau_i \lambda_{is} \right) \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \sum_{s=j+1}^p \nu_s \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T, \end{aligned}$$

a matriz  $\mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} - \nu_j \mathbf{I}_p = \sum_{s=j+1}^p \nu_s \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T - \nu_j \mathbf{I}_p$  tem característica completa e inversa

$$(\mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} - \nu_j \mathbf{I}_p)^{-1} = \sum_{s=j+1}^p \frac{1}{\nu_s - \nu_j} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T - \sum_{s=1}^j \frac{1}{\nu_j} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T.$$

Então vem,  $(\mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} - \nu_j \mathbf{I}_p)^{-1} \boldsymbol{\beta}_s = -\frac{1}{\nu_j} \boldsymbol{\beta}_s$  para  $s = 1, \dots, j-1$ , e

$$(\mathbf{P}_{j+1} \boldsymbol{\Sigma}_{\text{id}} - \nu_j \mathbf{I}_p)^{-1} \mathbf{P}_{j+1} = \sum_{s=j+1}^p \frac{1}{\nu_s - \nu_j} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T.$$

Assim, de (4.63) obtém-se

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= - \sum_{s=1}^{j-1} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s - \\ &- \frac{1}{2} \sum_{s=j+1}^p \frac{1}{\nu_s - \nu_j} \boldsymbol{\beta}_s \tau_i \sqrt{\lambda_{ij}} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \boldsymbol{\beta}_s^T \mathbf{x}. \end{aligned} \quad (4.64)$$

Do resultado (4.64) tem-se que para qualquer  $s = j+1, \dots, p$

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T \boldsymbol{\beta}_s = \frac{1}{2(\nu_j - \nu_s)} \boldsymbol{\beta}_s \tau_i \sqrt{\lambda_{ij}} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \boldsymbol{\beta}_s^T \mathbf{x}.$$

Por último, usando o facto de  $\text{IF}(y, \sigma^2, G_0) = 2\text{IF}(y, \sigma, G_0)$ , obtém-se (4.50).

- *Funções de influência parciais dos valores próprios.*

Usando o facto de  $\lambda_{ij,\varepsilon,i} = \sigma^2 (F_{i,\mathbf{x},\varepsilon} [\boldsymbol{\beta}_{j,\varepsilon,i}])$  para  $\ell \neq i$ , e de  $\lambda_{\ell j,\varepsilon,i} = \sigma^2 (F_\ell [\boldsymbol{\beta}_{j,\varepsilon,i}]) = \boldsymbol{\beta}_{j,\varepsilon,i}^T \boldsymbol{\Sigma}_\ell \boldsymbol{\beta}_{j,\varepsilon,i}$ , com a regra da diferenciação da função composta, resulta

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \lambda_{\sigma,ij}, F) &= \frac{\partial}{\partial \varepsilon} \sigma^2 (F_{i,\varepsilon,\mathbf{x}} [\boldsymbol{\beta}_j]) \big|_{\varepsilon=0} + \left[ \frac{\partial}{\partial \mathbf{b}} \sigma^2 (F_i [\mathbf{b}]) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} \right]^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i(\mathbf{b})) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} + (2\boldsymbol{\Sigma}_i \boldsymbol{\beta}_j)^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i(\mathbf{b})) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} + 2\lambda_{ij} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= 2\lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right), \end{aligned} \quad (4.65)$$

para  $i = 1, \dots, k$ ,  $j = 1, \dots, p$ .

Então, vem que  $\text{PIF}_i(\mathbf{x}, \lambda_{\sigma,\ell j}, F) = 2\lambda_{\ell j} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = 0$ , para  $i = 1, \dots, k$  e  $\ell \neq i$ , o que conclui a prova de (4.49).

- *Funções de influência parciais das matrizes de dispersão.*

Uma vez que  $\mathbf{V}_{\ell,\varepsilon,i} = \mathbf{V}_{\sigma,\ell}(F_{\varepsilon,\mathbf{x},i}) = \sum_{j=1}^p \lambda_{\ell j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i}^T$ , e usando as expressões das *funções de influência parciais* obtidas para os vectores e valores próprios, vem

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) &= \sum_{j=1}^p \text{PIF}_i(\mathbf{x}, \lambda_{\sigma,\ell j}, F) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \sum_{j=1}^p \lambda_{\ell j} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=1}^p \lambda_{\ell j} \boldsymbol{\beta}_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T. \end{aligned}$$

Assim,

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=2}^p \lambda_{\ell j} \sum_{s=1}^{j-1} \frac{1}{\nu_j - \nu_s} \tau_i \sqrt{\lambda_{is}} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) \times \\ &\times (\mathbf{x}^T \boldsymbol{\beta}_j) [\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T] + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \lambda_{\ell s} \frac{1}{\nu_s - \nu_j} \tau_i \sqrt{\lambda_{is}} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) \times \\ &\times (\mathbf{x}^T \boldsymbol{\beta}_j) [\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T], \end{aligned}$$

resultando finalmente que

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma, \ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\nu_j - \nu_s} \tau_i \sqrt{\lambda_{is}} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) \times \\ &\times (\mathbf{x}^T \boldsymbol{\beta}_j) [\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T] , \end{aligned}$$

o que conclui a prova.  $\square$

Analogamente ao exposto na Subsecção 3.1, as *funções de influência parciais* dos funcionais valores próprios são semelhantes às obtidas por Croux e Ruiz-Gazen (2000). Por outro lado, no que diz respeito às *funções de influência parciais* dos vectores próprios comuns, embora a forma destas funções seja a mesma da deduzida por Croux e Ruiz-Gazen (2000) para o modelo das *componentes principais*, nota-se a inclusão de pesos relacionados com os valores próprios da matriz de dispersão combinada,  $\boldsymbol{\Sigma}_{\text{id}}$ . Quando  $k = 1$  em (4.50) a *função de influência* do funcional vector próprio vem

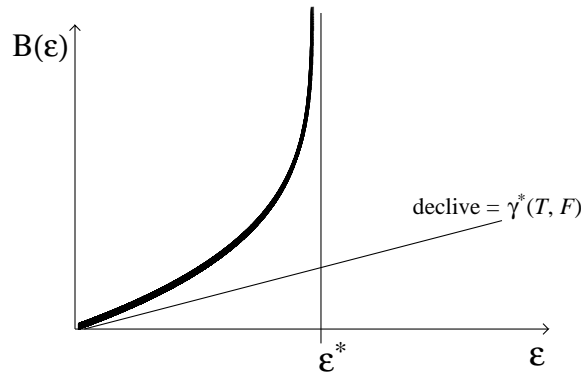
$$\begin{aligned} \text{IF}(\mathbf{x}, \boldsymbol{\beta}_{\sigma, j}, F) &= \sqrt{\lambda_j} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_j}}, \sigma, G_0 \right) \sum_{s=j+1}^p \frac{1}{\lambda_j - \lambda_s} \boldsymbol{\beta}_s (\mathbf{x}^T \boldsymbol{\beta}_s) + \\ &+ \sum_{s=1}^{j-1} \frac{1}{\lambda_j - \lambda_s} \boldsymbol{\beta}_s \sqrt{\lambda_s} \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_s}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) , \quad (4.66) \end{aligned}$$

coincidindo com a obtida por Croux e Ruiz-Gazen (2000), como não podia deixar de ser.

A principal alteração de (4.50) em relação a (4.66) é a modificação do denominador  $(\lambda_j - \lambda_s)$  para  $(\nu_j - \nu_s)$ . Isto é compreensível, uma vez que para o modelo das *componentes parciais comuns* este funcional é comum às  $k$  populações.

Observando as equações (4.49) e (4.50) é fácil concluir que a escolha de um funcional de escala com *função de influência* limitada origina que os valores próprios tenham *funções de influência parciais* limitadas. Por sua vez, para o funcional vectores próprios, as suas *funções de influência parciais* mantêm-se sempre ilimitadas, independentemente da escolha do funcional de escala. Isto deve-se ao facto de o crescimento do termo  $\mathbf{x}^T \boldsymbol{\beta}_s$ ,  $s = j + 1, \dots, p$ , não poder ser compensado por  $\text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_j}}, \sigma, G_0 \right)$  porque este actua em direcções ortogonais uma vez que  $j \neq s$ . Este problema também existe para o modelo das *componentes principais*, como se constata na equação (4.66).

Com este método de estimação é impossível construir um estimador *B-robusto* para os vectores próprios comuns. No entanto, embora os estimadores dos vectores próprios sejam sensíveis a pequenas quantidades de contaminação, podem possuir boas propriedades de robustez global. Croux e Ruiz-Gazen (2000) referem esta característica dos estimadores PP, no contexto da análise em *componentes principais*. Por exemplo, para possuírem um *alto ponto de ruptura* basta considerar um estimador de escala com esta característica. Em complemento das *funções de influência parciais*, que avaliam localmente a robustez dos funcionais, as *curvas de enviesamento máximo parciais* (PB) indicam a robustez global dos funcionais. Estas curvas, que generalizam a *curva de enviesamento máximo* (“maxbias curve”, Huber, 1964) para  $k$  populações, são definidas da forma que se segue.



**Figura 4.3:** Curva de enviesamento máximo de um estimador genérico,  $T$ , com *ponto de ruptura* positivo e *função de influência* limitada.

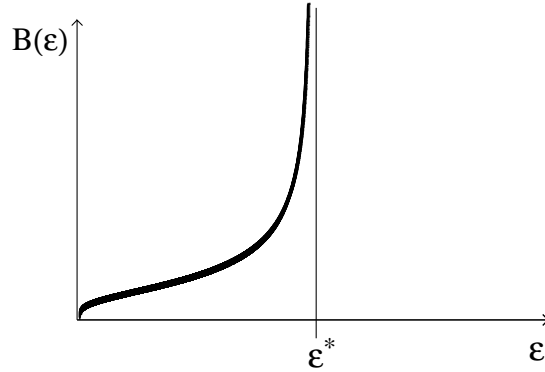
**Definição 4.5 (curvas de enviesamento máximo parciais)** Seja  $F$  a medida produto,  $F = F_1 \times \cdots \times F_k$ . Considere-se  $GE_{\epsilon, i_0} = F_1 \times \cdots \times F_{i_0-1} \times GE_{i_0, \epsilon} \times F_{i_0+1} \times \cdots \times F_k$  com  $GE_{i, \epsilon} = GE(F_i, \epsilon) = \{G_i : G_i = (1 - \epsilon)F_i + \epsilon H_i, H_i \in \mathcal{F}\}$ .

As *curvas de enviesamento máximo parciais* (PB) do funcional  $\mathbf{T}$ ,  $p$ -dimensional, em  $F$  são dadas por

$$PB_{i_0}(\epsilon) = PB_{i_0}(\epsilon, \mathbf{T}, F) = \sup_{G_{i_0} \in GE_{i_0, \epsilon}} \|\mathbf{T}(GE_{\epsilon, i_0}) - \mathbf{T}(F)\|, \quad i_0 = 1, \dots, k. \quad (4.67)$$

A *curva de enviesamento máximo* é bastante útil e sumaria várias medidas da robustez dos funcionais, incluindo o seu *ponto de ruptura* e a *sensibilidade a grandes*





**Figura 4.4:** Curva de enviesamento máximo de um estimador genérico,  $T$ , com *ponto de ruptura* positivo e *função de influência* ilimitada.

*erros*. Quando a *curva de enviesamento máximo* é contínua em  $\varepsilon = 0$ , então  $\mathbf{T}$  considera-se tanto mais robusto em  $F$ , em sentido quantitativo, quanto mais rápida for a convergência desta curva para zero com  $\varepsilon$  (sob condições de regularidade tem-se que  $\frac{\partial}{\partial \varepsilon} \text{PB}_i(\varepsilon) = \gamma_i^*$ ). Regra geral a *curva de enviesamento máximo* dá uma descrição mais completa da robustez do funcional que ambas as características referidas anteriormente, o *ponto de ruptura* e a *sensibilidade a grandes erros*. As Figuras 4.3 e 4.4 ilustram qualitativamente, para um funcional genérico, o que acabamos de explicar. No entanto, como referem Croux e Haesbroeck (2002), a construção das *curvas de enviesamento máximo* geralmente não é um tema simples de abordar, especialmente em situações multivariadas.

Voltando à escolha do funcional de escala  $\sigma$ , é simples verificar que quando este é seleccionado de forma a ser equivalente ao estimador variância, ambas as *funções de influência parciais* se tornam ilimitadas. Considerando  $\sigma^2(F) = \text{VAR}(F)$  e como

$$\text{IF}(y, \sigma, G) = \frac{1}{2} \{y^2 - \text{VAR}(G)\},$$

resulta que

$$\text{PIF}_i(\mathbf{x}, \lambda_{\text{VAR}, \ell_j}, F) = \delta_{\ell i} \left\{ (\mathbf{x}^T \boldsymbol{\beta}_j)^2 - \lambda_{ij} \right\}, \quad (4.68)$$

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\text{VAR}, j}, F) = \tau_i \sum_{s \neq j}^p \frac{1}{\nu_j - \nu_s} \mathbf{x}^T \boldsymbol{\beta}_s \mathbf{x}^T \boldsymbol{\beta}_j \boldsymbol{\beta}_s. \quad (4.69)$$

Os autores Croux e Ruiz-Gazen (2000) apresentaram gráficos das *funções de influência* para os valores e vectores próprios quando  $\sigma$  é o desvio padrão e também

quando  $\sigma$  é o funcional robusto relacionado com o estimador-Q (Rousseeuw e Croux, 1992, 1993). Este estimador com *alto ponto de ruptura* é dado pelo primeiro quartil da amostra das diferenças entre pares de observações, i.e.,

$$Q = d \{ |X_i - X_j| ; 1 \leq i < j \leq n \}_{(k)},$$

onde  $d$  é um factor para obter consistência e  $k = \binom{[n/2]+1}{2}$ .

Quanto às variâncias assintóticas dos estimadores PP dos vectores próprios comuns e dos valores próprios, estas podem ser obtidas recorrendo a (4.6).

**Corolário 4.2 (variâncias assintóticas dos estimadores PP do modelo CPC)**

Sejam  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i})$  com  $i = 1, \dots, k$ , vectores aleatórios independentes de  $k$  amostras, supostamente também independentes, e com distribuição  $F_i$ . Suponha-se que a localização é  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e a matriz de dispersão  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$  é tal que  $\boldsymbol{\Sigma}_i$  satisfaz (2.3) com  $\boldsymbol{\beta} = \mathbf{I}_p$ . Assuma-se que  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e que  $\sum_{i=1}^k \tau_i = 1$ .

Considere-se o vector  $\boldsymbol{\Lambda}_i^{-\frac{1}{2}} \mathbf{X}_{i1} = \mathbf{Z}_i$ , com a mesma distribuição esférica  $G$  para todo  $i = 1, \dots, k$ . Assuma-se ainda que,  $\sigma(G_0) = 1$  onde  $G_0$  é a distribuição de  $Z_{11}$ , e que  $\boldsymbol{\Sigma}_{\text{id}} = \text{diag}(\nu_1, \dots, \nu_p)$ , com  $\nu_1 > \nu_2 > \dots > \nu_p$ .

Representando por  $s(\cdot)$  um estimador univariado relacionado com  $\sigma(\cdot)$ , os estimadores dos eixos principais comuns definem-se resolvendo iterativamente

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= \arg \max_{\|\mathbf{b}\|=1} \sum_{i=1}^k \tau_i s^2(\mathbf{X}_i^T \mathbf{b}), \\ \hat{\boldsymbol{\beta}}_j &= \arg \max_{\mathbf{b} \in \mathcal{B}_j} \sum_{i=1}^k \tau_i s^2(\mathbf{X}_i^T \mathbf{b}), \quad j = 2, \dots, p, \end{aligned} \quad (4.70)$$

onde  $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}^T \hat{\boldsymbol{\beta}}_m = 0 \text{ para } m = \dots, j-1\}$ .

Os estimadores dos valores próprios e da matriz de dispersão da  $i$ -ésima população são calculados por

$$\hat{\lambda}_{ij} = s^2(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_j) \quad \text{para } j = 1, \dots, p, \quad \mathbf{V}_i = \sum_{j=1}^p \hat{\lambda}_{ij} \hat{\boldsymbol{\beta}}_j \hat{\boldsymbol{\beta}}_j^T. \quad (4.71)$$

Então, no caso da função  $(\varepsilon, y) \rightarrow \sigma\{(1-\varepsilon)G_0 + \varepsilon\Delta_y\}$  ser contínua e duas vezes diferenciável em  $(0, y)$ , as variâncias assintóticas dos estimadores PP, solução de

(4.70), são dadas por

$$\begin{aligned}\text{ASVAR}(\widehat{\lambda}_{\ell j}) &= 4 \lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(s, G_0) , \\ \text{ASVAR}(\widehat{\beta}_{jm}) &= \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 , \quad \text{para } m \neq j , \\ \text{ASVAR}(\widehat{\beta}_{jm}) &= 0 , \quad \text{para } m = j .\end{aligned}$$

Em particular, quando  $G = N(\mathbf{0}_p, \mathbf{I}_p)$ , tem-se

$$\begin{aligned}\text{ASVAR}(\widehat{\lambda}_{\ell j}) &= 4 \lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(s, \Phi) , \\ \text{ASVAR}(\widehat{\beta}_{jm}) &= \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2} E_\Phi \{ \text{DIF}(Y, \sigma, \Phi) \}^2 , \quad \text{para } m \neq j , \\ \text{ASCOV}(\widehat{\beta}_{jm}, \widehat{\beta}_{jr}) &= 0 , \quad \text{para } m \neq j , \quad m \neq r , \quad r \neq j .\end{aligned}$$

**Dem:** De (4.6), usando-se (4.49) e o facto de que  $\frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\lambda_{\ell j}^{\frac{1}{2}}} \sim G_0$  quando  $\mathbf{X} \sim F_i$ , obtém-se

$$\text{ASVAR}(\widehat{\lambda}_{\ell j}) = \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} \left[ \left\{ \delta_{\ell i} 2 \lambda_{ij} \text{IF} \left( \frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\lambda_{ij}^{\frac{1}{2}}}, \sigma, G_0 \right) \right\}^2 \right] = 4 \lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(s, G_0) .$$

De (4.50) e uma vez que  $\boldsymbol{\beta} = \mathbf{I}_p$ , obtém-se

$$\begin{aligned}\text{ASVAR}(\widehat{\beta}_{jm}) &= \delta_{m>j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 + \\ &+ \delta_{m<j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2} E_G \{ \text{DIF}(Z_{1m}, \sigma, G_0) Z_{1j} \}^2 = \\ &= \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 ,\end{aligned}$$

onde  $\delta_{m>j} = 0$  se  $m \leq j$  e  $\delta_{m>j} = 1$  se  $m > j$ .

Para a covariância assíptótica, sabemos que

$$\text{ASCOV}(\widehat{\beta}_{jm}, \widehat{\beta}_{jr}) = \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} [ \text{PIF}_i(\mathbf{X}, \boldsymbol{\beta}_{\mathbf{V}, jm}, F) \text{PIF}_i(\mathbf{X}, \boldsymbol{\beta}_{\mathbf{V}, jr}, F) ] .$$

Assim, quando  $G = N(\mathbf{0}_p, \mathbf{I}_p)$ , e uma vez que  $Z_{1j}$ ,  $Z_{1m}$  and  $Z_{1r}$  são independentes

para  $j \neq r$ ,  $j \neq m$  e  $m \neq r$  e  $E_{\Phi} \{ \text{DIF} (Y, \sigma, \Phi) \} = 0$ , é fácil verificar que

$$\begin{aligned}
\text{ASCOV} \left( \widehat{\beta}_{jm}, \widehat{\beta}_{jr} \right) &= \delta_{m>j} \delta_{r>j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} (\lambda_{im} \lambda_{ir})^{\frac{1}{2}}}{(\nu_j - \nu_m)(\nu_j - \nu_r)} \times \\
&\times E_G \left[ \{ \text{DIF} (Z_{1j}, \sigma, G_0) \}^2 Z_{1m} Z_{1r} \right] + \\
&+ \delta_{m<j} \delta_{r<j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} (\lambda_{im} \lambda_{ir})^{\frac{1}{2}}}{(\nu_j - \nu_m)(\nu_j - \nu_r)} \times \\
&\times E_G \left\{ \text{DIF} (Z_{1m}, \sigma, G_0) \text{DIF} (Z_{1r}, \sigma, G_0) (Z_{1j})^2 \right\} + \\
&+ \delta_{m>j} \delta_{r<j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} (\lambda_{im} \lambda_{ir})^{\frac{1}{2}}}{(\nu_j - \nu_m)(\nu_j - \nu_r)} \times \\
&\times E_G \left\{ \text{DIF} (Z_{1j}, \sigma, G_0) \text{DIF} (Z_{1r}, \sigma, G_0) Z_{1m} Z_{1j} \right\} + \\
&+ \delta_{m<j} \delta_{r>j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} (\lambda_{im} \lambda_{ir})^{\frac{1}{2}}}{(\nu_j - \nu_m)(\nu_j - \nu_r)} \times \\
&\times E_G \left\{ \text{DIF} (Z_{1j}, \sigma, G_0) \text{DIF} (Z_{1m}, \sigma, G_0) Z_{1r} Z_{1j} \right\},
\end{aligned}$$

resultando então  $\text{ASCOV} \left( \widehat{\beta}_{jm}, \widehat{\beta}_{jr} \right) = 0$ , o que conclui a prova.  $\square$

É fácil concluirmos que, para o caso particular do funcional  $\sigma^2(F) = \text{VAR}(F)$ , a variância assintótica dos estimadores dos valores próprios é igual à dos estimadores de máxima verosimilhança. Por outro lado, no que se refere à variância assintótica do estimador dos vectores próprios, esta não iguala a variância assintótica dos estimadores de máxima verosimilhança. Neste caso, a variância assintótica é

$$\text{ASVAR} \left( \widehat{\beta}_{jm} \right) = \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \lambda_{im}}{(\nu_j - \nu_m)^2},$$

alcançando-se a variância assintótica dos estimadores construídos à custa da matriz de dispersão combinada, Boente e Orellana (2001). Note-se que a eficiência assintótica destes estimadores depende dos valores próprios de cada  $\Sigma_i$  e também de  $\Sigma_{\text{id}}$ .

As *funções de influência parciais* aqui construídas vão-nos permitir propor técnicas para a detecção de observações discordantes tanto para o caso das *componentes principais*, como para as *componentes principais comuns*. Todo este estudo será apresentado em pormenor no Capítulo 6 desta tese. Além disso, no capítulo que se segue serão introduzidos novos estimadores “projection-pursuit” e será feito o estudo das suas propriedades de robustez via *funções de influência parciais*.



## Capítulo 5

# Estimadores “Projection–Pursuit” Gerais

### 5.1 Introdução

Neste capítulo pretendemos generalizar os estimadores de “projection–pursuit” apresentados anteriormente. Esta decisão foi tomada face a algumas vantagens destes estimadores em relação à alternativa “plug–in”. Embora se possa construir o processo de “plug–in” com estimadores com *alto ponto de ruptura*, por exemplo empregando o MCD (Rousseeuw, 1984) como Croux e Haesbroeck (2000) fizeram na estimação das *componentes principais*, esta metodologia vê-se limitada a problemas com dimensões pequenas ou moderadas. Para se ver porquê, considere-se a técnica “plug–in” com o MCD. Este estimador, como se viu, é definido à custa da média e da matriz de covariâncias do subconjunto das  $h$  observações com o menor determinante da matriz de covariâncias. Mas, o MCD só pode ser calculado se o número de variáveis for  $p < h$ , pois caso contrário a matriz de covariâncias de qualquer subconjunto  $h$  tem determinante zero. Além disso, em todas as situações a alternativa “plug–in” só é aplicável a conjuntos de dados com mais observações por grupo do que variáveis. Por vezes podem ocorrer situações onde isto não se verifica e nesse caso o método PP é uma alternativa. Por outro lado, para o caso particular do modelo CPC, Boente e Orellana (2001) concluem, com base no estudo de Monte Carlo por elas desenvolvido, que os estimadores “plug–in” mostram uma grande sensibilidade a contaminações na direcção do menor valor próprio, sendo assim preferível a alternativa PP.

Uma outra vantagem da alternativa PP advém do seu processo sequencial de estimação. Este facto permite parar a estimação em qualquer passo, não sendo

necessário esperar pela estimação de todas as direcções comuns. Isto é particularmente atractivo quando o objectivo do estudo recai sobre um subconjunto dos vectores próprios comuns, associados com maior (menor) variabilidade. Com esta técnica consegue-se economizar tempo e recursos relativamente ao processo “plug-in”.

Dadas as vantagens do método PP pareceu-nos interessante tentar construir novos estimadores com esta técnica. Seria muito vantajoso dispormos de novos procedimentos com melhores desempenhos na estimação dos parâmetros do modelo CPC. Em especial, seria interessante criar estimadores que alcançassem melhores resultados na estimação das direcções principais comuns. Isto porque como se pode ver, na simulação desenvolvida por Boente e Orellana (2001), os estimadores PP das direcções principais em estudo mostraram-se ainda sensíveis à contaminação exercida (embora menos que os “plug-in”).

Com este objectivo em mente estruturámos este capítulo da seguinte forma. Na Secção 2 apresentamos os novos estimadores e a motivação da sua definição. O estudo da consistência segundo Fisher destes estimadores encontra-se na Secção 3. De forma a poder avaliar a robustez das novas propostas construímos na Secção 4 as *funções de influência parciais* dos funcionais introduzidos. Apresentamos também a derivação heurística das variâncias assintóticas dos estimadores relacionados com estes funcionais. A investigação desenvolvida conduziu ao estudo particular dos estimadores com a aplicação da função logaritmo, tendo sido possível demonstrar a optimalidade desses estimadores para o caso particular do modelo proporcional. Todos estes resultados são apresentados na Secção 5.

Por último, Secção 6 e Secção 7, comparamos o desempenho dos estimadores aqui propostos com os sugeridos por Boente e Orellana (2001). Para isso consideramos em primeiro lugar três situações de dados simulados com amostras normalmente distribuídas e amostras contaminadas. Em segundo lugar são estudados os dados reais do Exemplo 1 (Hemofilia) da Subsecção 3.3 do Capítulo 2 e também um outro conjunto de dados reais, Exemplo 2 (Variedades de Castanheiros), composto por 2 grupos num espaço de dimensão 5. Refira-se ainda que a comparação com a abordagem de Novi Inverardi e Flury (1992) e de Orellana (1999) está também subjacente ao nosso estudo, uma vez que o trabalho de Boente e Orellana (2001) contemplou esta questão. As conclusões mais importantes desta pesquisa foram condensadas no artigo Boente, Pires e Rodrigues (2002b), submetido para publicação em Julho de 2002.

## 5.2 Definição dos estimadores

No Capítulo 4 abordámos os estimadores PP das direcções principais comuns e dispersões, introduzidos por Boente e Orellana (2001). Estes estimadores, apresentados em (4.46), constroem-se à custa da maximização (ou minimização) de uma medida de dispersão combinada das observações projectadas.

Por outro lado, como já referimos, seguindo Flury e Gautschi (1986) e Flury (1988, pág. 68) os estimadores de máxima verosimilhança de  $\beta$  também minimizam

$$\prod_{i=1}^k \left[ \frac{\det \{\text{diag}(\mathbf{F}_i)\}}{\det(\mathbf{F}_i)} \right]^{n_i}, \quad (5.1)$$

onde  $\mathbf{F}_i = \beta^T \mathbf{S}_i \beta$ , e  $\mathbf{S}_i$  é a matriz de covariâncias amostral da  $i$ -ésima população. Novamente  $(n_i - 1)$  foi substituído por  $n_i$  de forma a garantir a equivalência entre os funcionais e os estimadores.

Por sua vez, a minimização de (5.1) é também equivalente a minimizar

$$\begin{aligned} \ln \left( \prod_{i=1}^k \left[ \frac{\det \{\text{diag}(\mathbf{F}_i)\}}{\det(\mathbf{F}_i)} \right]^{n_i} \right) &= \sum_{i=1}^k n_i \ln [\det \{\text{diag}(\mathbf{F}_i)\}] - \sum_{i=1}^k n_i \ln \{\det(\mathbf{F}_i)\} = \\ &= \sum_{i=1}^k n_i \ln [\det \{\text{diag}(\mathbf{F}_i)\}] - \sum_{i=1}^k n_i \ln \{\det(\mathbf{S}_i)\} = \\ &= \sum_{i=1}^k n_i \sum_{j=1}^p \ln(\ell_{ij}) - \sum_{i=1}^k n_i \ln \{\det(\mathbf{S}_i)\}, \end{aligned}$$

onde  $\ell_{ij}$  são os elementos diagonais de  $\mathbf{F}_i$ . Portanto, o estimador de máxima verosimilhança  $\hat{\beta}$  pode ser visto como a solução minimizante de  $\sum_{i=1}^k n_i \sum_{j=1}^p \ln(\ell_{ij})$ . Por outro lado, tendo em conta que  $\hat{\lambda}_{ij} = \hat{\beta}_j^T \mathbf{S}_i \hat{\beta}_j$  iguala a variância amostral das variáveis projectadas, um novo procedimento “projection-pursuit” robusto pode ser obtido resolvendo iterativamente

$$\hat{\beta}_1 = \arg \max_{\|\mathbf{b}\|=1} \sum_{i=1}^k \tau_i \ln \{s^2(\mathbf{X}_i^T \mathbf{b})\}, \quad \hat{\beta}_j = \arg \max_{\mathbf{b} \in \mathcal{B}_j} \sum_{i=1}^k \tau_i \ln \{s^2(\mathbf{X}_i^T \mathbf{b})\}, \quad j = 2, \dots, p, \quad (5.2)$$

onde  $s$  é um estimador de escala robusto.

Outra solução seria obtida tomando o operador máximo, analogamente ao que acontece no caso de apenas uma população.

Como vimos, a pré-aplicação do operador logaritmo em (5.2) à variabilidade das variáveis projectadas surgiu em consequência da solução de máxima verosimilhança.



Contudo, do mesmo modo que aplicamos a função logaritmo pode-se utilizar qualquer outra função, estritamente crescente e positiva, e com ela criar novos estimadores PP robustos.

Assim, um procedimento PP mais geral será obtido considerando uma função estritamente crescente  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  e as direcções comuns estimadas como

$$\hat{\boldsymbol{\beta}}_1 = \arg \max_{\|\mathbf{b}\|=1} \sum_{i=1}^k \tau_i f \{s^2(\mathbf{X}_i^T \mathbf{b})\}, \quad \hat{\boldsymbol{\beta}}_j = \arg \max_{\mathbf{b} \in \mathcal{B}_j} \sum_{i=1}^k \tau_i f \{s^2(\mathbf{X}_i^T \mathbf{b})\}, \quad j = 2, \dots, p. \quad (5.3)$$

Considerando o operador mínimo em vez de máximo surgirá uma definição diferente de (5.3) mas ambos os casos conduzem às mesmas *funções de influência parciais*, pelo que são assintoticamente equivalentes.

Quanto aos estimadores dos valores próprios da  $i$ -ésima população, estes podem ser calculados como  $\hat{\lambda}_{ij} = s^2(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_j)$  para  $j = 1, \dots, p$ . Para evitar ambiguidades os vectores próprios serão posteriormente ordenados por ordem decrescente dos valores próprios da primeira população. Havendo casos de igualdade estre ordenam-se pela segunda população e assim sucessivamente. (Note-se que o modelo CPC exige que para cada par de valores próprios exista pelo menos uma população onde eles são distintos.)

### 5.3 Fisher-consistência

Para apresentar as condições segundo as quais os funcionais em estudo são Fisher-consistentes para os parâmetros de interesse considera-se que  $\mathbf{X}_i$ ,  $i = 1, \dots, k$ , representam vectores aleatórios independentes e identicamente distribuídos com  $F_i$ . A distribuição  $F_i$  tem parâmetro de localização  $\boldsymbol{\mu}_i$  e matriz de dispersão  $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^T$  que satisfaz (2.3). Analogamente ao que foi feito no capítulo anterior, e sem perda de generalidade, vamos assumir que  $\boldsymbol{\mu}_i = \mathbf{0}_p$ . Representa-se por  $F_i[\mathbf{b}]$  a distribuição de  $\mathbf{b}^T \mathbf{X}_i$  e por  $F$  a medida produto definida como  $F = F_1 \times \dots \times F_k$ .

Seja  $\varsigma(\mathbf{b}) = \sum_{i=1}^k \tau_i f\{\sigma^2(F_i[\mathbf{b}])\}$ , onde  $\sigma(\cdot)$  indica o funcional de escala univariado relacionado com o estimador  $s$ . Admite-se que o estimador de escala  $s$  é *equivariante* face a transformações de escala.

Neste contexto, os funcionais “projection-pursuit” das direcções comuns  $\boldsymbol{\beta}_\sigma(F) =$

$(\beta_{\sigma,1}(F), \dots, \beta_{\sigma,p}(F))$  são solução de

$$\begin{aligned}\beta_{\sigma,1}(F) &= \arg \max_{\|\mathbf{b}\|=1} \varsigma(\mathbf{b}) , \\ \beta_{\sigma,j}(F) &= \arg \max_{\mathbf{b} \in \mathcal{B}_j} \varsigma(\mathbf{b}) , \quad j = 2, \dots, p ,\end{aligned}\tag{5.4}$$

onde  $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}^T \beta_m(F) = 0, \text{ para } m = 1, \dots, j-1\}$ , segundo a ordenação proposta no caso dos estimadores, enquanto que os funcionais valores próprios e matrizes de dispersão são definidos como

$$\lambda_{\sigma,ij}(F) = \sigma^2 (F_i [\beta_{\sigma,j}(F)]) ,\tag{5.5}$$

$$\mathbf{V}_{\sigma,i}(F) = \sum_{j=1}^p \lambda_{\sigma,ij}(F) \beta_{\sigma,j}(F) \beta_{\sigma,j}(F)^T .\tag{5.6}$$

Boente e Orellana (2001) obtiveram as condições sob as quais os funcionais definidos por (5.4) serão Fisher-consistentes, no caso de  $f(t) = \text{id}(t) = t$ .

O teorema seguinte estabelece a Fisher-consistência dos funcionais (5.4) e (5.5), para uma função geral  $f$  e, também, para o caso especial de  $f(t) = \ln(t)$ .

### **Teorema 5.1 (Fisher-consistência dos funcionais PP gerais)**

Sejam  $\mathbf{X}_i$ ,  $i = 1, \dots, k$ , vectores aleatórios independentes com distribuição elipticamente simétrica  $F_i$ , com parâmetros de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matrizes de dispersão  $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^T$  que satisfazem (2.3). Assuma-se que  $\mathbf{C}_i^{-1} \mathbf{X}_i = \mathbf{Z}_i$  tem a mesma distribuição esférica  $G$  para todo  $i = 1, \dots, k$ . Denote-se por  $G_0$  a distribuição de  $Z_{11}$  e por  $F_i[\mathbf{b}]$  a distribuição de  $\mathbf{b}^T \mathbf{X}_i$ .

Seja  $\sigma(\cdot)$  um funcional de escala, *equivariante* para transformações de escala, e tal que  $\sigma(G_0) = 1$ . Represente-se por  $\beta_\sigma(F) = (\beta_{\sigma,1}(F), \dots, \beta_{\sigma,p}(F))$  a solução de (5.4) e seja  $\lambda_{\sigma,ij}(F) = \sigma^2 (F_i[\beta_{\sigma,j}(F)])$ .

Então, tem-se que

(i) se  $f$  é estritamente crescente e diferenciável,  $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$  é uma matriz com  $\lambda_{i1} \geq \dots \geq \lambda_{ip}$  para  $i = 1, \dots, k$  e se existir um índice  $i_0 = i_0(j)$  tal que  $\lambda_{i_0 j} > \lambda_{i_0, j+1}$  para  $j = 1, \dots, p-1$ , então os funcionais  $\beta_\sigma(F)$  e  $\lambda_{\sigma,ij}(F)$  são Fisher-consistentes;

(ii) se  $f$  é estritamente crescente de classe  $C^1$  e  $f'$  não constante e se  $\exists_{j \neq m}, (\nu_j - \nu_{jm})(\nu_{mj} - \nu_m) < 0$ , onde  $\nu_{js} = \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{is}$  e  $\nu_j = \nu_{jj}$ , então os funcionais não são Fisher-consistentes;

(iii) se  $f$  é estritamente crescente, de classe  $C^1$ , e  $f''(t) < 0$ ,  $\forall t > 0$  ( $f''(t) > 0$ ,  $\forall t > 0$ ), então os funcionais obtidos por minimização (maximização) são Fisher-consistentes se e só se para cada par de valores próprios existir pelo menos uma população para a qual eles são distintos, i.e.,  $\forall_{j \neq m} \exists_{i_0} : \lambda_{i_0 j} \neq \lambda_{i_0 m}$ ;

(iv) se  $f$  é estritamente crescente, de classe  $C^1$ , e  $f''(t) < 0$ ,  $\forall t > 0$  ( $f''(t) > 0$ ,  $\forall t > 0$ ), então os funcionais obtidos por maximização (minimização) são Fisher-consistentes se e só se  $\forall_{j \neq m} \exists_{i_0} : \lambda_{i_0 j} \neq \lambda_{i_0 m}$  e  $\forall_{j \neq m} (\nu_j - \nu_{jm})(\nu_{mj} - \nu_m) \geq 0$ .

**Dem:** Sem perda de generalidade considere-se o caso diagonal, i.e.,  $\Sigma_i = \Lambda_i$ . Neste caso, a condição de Fisher-consistência para os vectores próprios é equivalente a  $\beta_\sigma(F) = \mathbf{I}_p$ .

Vamos então supor que se pretende determinar o primeiro vector próprio  $\beta_1$  que maximiza  $\varsigma(F[\mathbf{b}])$  sujeito à condição  $\|\beta_1\| = 1$ . Assim,  $\beta_1$  maximiza (minimiza) a função Lagrangeana

$$L(\mathbf{b}, \gamma) = \sum_{i=1}^k \tau_i f(\mathbf{b}^T \Lambda_i \mathbf{b}) - \gamma (\mathbf{b}^T \mathbf{b} - 1) , \quad (5.7)$$

e tem que satisfazer a equação

$$\mathbf{0}_p = \frac{\partial}{\partial \mathbf{b}} L(\mathbf{b}, \gamma)|_{\mathbf{b}=\beta_1} = 2 \sum_{i=1}^k \tau_i \Lambda_i \beta_1 f'(\beta_1^T \Lambda_i \beta_1) - 2\gamma \beta_1 . \quad (5.8)$$

Admita-se novamente sem perda de generalidade que  $f'$  é não constante, pois o caso de  $f'$  constante (ou seja  $f(t) = \text{id}(t) \times \text{constante}$ ) foi tratado por Boente e Orellana (2001). É fácil verificar que para qualquer  $f$ , com  $f'$  não constante, todos os vectores da base canónica são solução de (5.8) sem qualquer outra condição adicional. A falta de consistência é equivalente à existência de outras soluções diferentes da base canónica. Para isto acontecer então pelo menos duas linhas da equação matricial (5.8) devem ser do tipo

$$\frac{\partial}{\partial b_j} L(\mathbf{b}, \gamma)|_{b_j=\beta_{1j}} = \sum_{i=1}^k \tau_i \lambda_{ij} \beta_{1j} f'(\beta_1^T \Lambda_i \beta_1) - \gamma \beta_{1j} = 0 , \quad (5.9)$$

com  $\beta_{1j} \neq 0$ , ou seja, pelo menos um par de equações

$$\left\{ \begin{array}{l} \sum_{i=1}^k \tau_i \lambda_{ij} f'(\beta_1^T \Lambda_i \beta_1) - \gamma = 0 , \\ \sum_{i=1}^k \tau_i \lambda_{im} f'(\beta_1^T \Lambda_i \beta_1) - \gamma = 0 . \end{array} \right. \quad (5.10)$$

O facto de  $\gamma$  ser único implica que

$$\sum_{i=1}^k \tau_i (\lambda_{ij} - \lambda_{im}) f'(\boldsymbol{\beta}_1^T \boldsymbol{\Lambda}_i \boldsymbol{\beta}_1) = 0. \quad (5.11)$$

(i) Uma vez que  $f'$  nunca se anula e os termos  $(\lambda_{ij} - \lambda_{im})$  têm todos o mesmo sinal e não são simultaneamente nulos, pela condição de teorema, conclui-se que a condição (5.11) é impossível, ou seja não existe solução para  $\boldsymbol{\beta}_1$  fora da base canónica.

Considerando agora a direcção  $\boldsymbol{\beta}_r$ , com  $r = 2, \dots, p$ , a equação Lagrangeana com a adição da restrição de ortogonalidade vem

$$L(\mathbf{b}, \gamma, \boldsymbol{\alpha}) = \sum_{i=1}^k \tau_i f(\mathbf{b}^T \boldsymbol{\Lambda}_i \mathbf{b}) - \gamma (\mathbf{b}^T \mathbf{b} - 1) - \sum_{s=1}^{r-1} \alpha_s \mathbf{b}^T \boldsymbol{\beta}_s. \quad (5.12)$$

Portanto, o vector  $\boldsymbol{\beta}_r$  tem que satisfazer a equação

$$\mathbf{0}_p = \frac{\partial}{\partial \mathbf{b}} L(\mathbf{b}, \gamma, \boldsymbol{\alpha})|_{\mathbf{b}=\boldsymbol{\beta}_r} = 2 \sum_{i=1}^k \tau_i \boldsymbol{\Lambda}_i \boldsymbol{\beta}_r f'(\boldsymbol{\beta}_r^T \boldsymbol{\Lambda}_i \boldsymbol{\beta}_r) - 2\gamma \boldsymbol{\beta}_r - \sum_{s=1}^{r-1} \alpha_s \boldsymbol{\beta}_s, \quad (5.13)$$

e as linhas desta equação matricial são do tipo

$$2 \sum_{i=1}^k \tau_i \lambda_{ij} \beta_{rj} f'(\boldsymbol{\beta}_r^T \boldsymbol{\Lambda}_i \boldsymbol{\beta}_r) - 2\gamma \beta_{rj} - \sum_{s=1}^{r-1} \alpha_s \beta_{sj} = 0, \quad j = 1, \dots, p. \quad (5.14)$$

Para  $r = 2$  admita-se que  $\boldsymbol{\beta}_1 = \mathbf{e}_1$  (primeiro vector da base canónica de  $\mathbb{R}^p$ ), pois no caso contrário já se verificou a não consistência. Isto implica que  $\beta_{21} = 0$  e que  $\alpha_1 = 0$  pelo que a parcela referente à ortogonalidade desaparece. Aplicando sucessivamente este argumento garante-se que as conclusões com base na análise do sistema (5.10) são válidas para  $r = 1, \dots, p$ , conduzindo portanto à demonstração da condição (i) para os vectores próprios. A Fisher-consistência relativamente aos valores próprios decorre imediatamente de  $\sigma(G_0) = 1$ .

(ii) Pelo argumento apresentado no final da demonstração de (i), relativo às restrições de ortogonalidade, basta analisar a determinação da primeira direcção principal comum. Por uma questão de simplicidade de notação, considera-se  $\mathbf{b}_1 = (b_1, \dots, b_p)^T$ . O problema em causa resume-se a encontrar os extremos da função

$$\xi(\mathbf{b}) = \sum_{i=1}^k \tau_i f(b_1^2 \lambda_{i1} + \dots + b_p^2 \lambda_{ip}), \quad (5.15)$$

sujeito a  $b_1^2 + \dots + b_p^2 = 1$ . Este problema é equivalente a determinar os extremos da função

$$\xi(\mathbf{z}) = \sum_{i=1}^k \tau_i f((\lambda_{i1} - \lambda_{ip}) z_1 + \dots + (\lambda_{i(p-1)} - \lambda_{ip}) z_p + \lambda_{ip}), \quad (5.16)$$

sujeito a  $0 \leq z_i \leq 1, i = 1, \dots, p-1$  e  $0 \leq 1 - z_1 - \dots - z_{p-1} \leq 1$ , o que é equivalente a exigir que  $\mathbf{z}$  pertença ao “simplex” de dimensão  $p-1$ ,  $\mathcal{S}_{p-1}$ . De notar que como  $\mathcal{S}_{p-1}$  é compacto e  $f$  é contínua existem extremos globais.

Seja  $\xi_{jm}(\mathbf{z})$  a restrição de  $\xi(\mathbf{z})$  sobre a fronteira que une os vértices correspondentes a  $z_j = 1$  e a  $z_m = 1$ ,

$$\xi_{jm}(\mathbf{z}) = \sum_{i=1}^k \tau_i f((\lambda_{ij} - \lambda_{ip}) z_j + \lambda_{im}) = \xi_{jm}(z_j) . \quad (5.17)$$

Tem-se assim que

$$\xi'_{jm}(z_j) = \sum_{i=1}^k \tau_i (\lambda_{ij} - \lambda_{im}) f'((\lambda_{ij} - \lambda_{im}) z_j + \lambda_{im}) . \quad (5.18)$$

Vamos agora supor que  $\exists_{j \neq m} (\nu_j - \nu_{jm})(\nu_{mj} - \nu_m) < 0$ , isto é,

$$\left[ \sum_{i=1}^k \tau_i (\lambda_{ij} - \lambda_{im}) f'(\lambda_{im}) \right] \left[ \sum_{i=1}^k \tau_i (\lambda_{ij} - \lambda_{im}) f'(\lambda_{ij}) \right] < 0 . \quad (5.19)$$

Esta condição implica, em face da continuidade de  $f'$ , que existe uma solução de  $\xi'_{jm}(z_j) = 0$  com  $0 < z_j < 1$ , pois  $\xi'_{jm}(0) \times \xi'_{jm}(1) < 0$ , o que corresponde a um vector não pertencente à base canónica e demonstra **(ii)**.

**(iii)** É fácil verificar que se  $f''(t) < 0, \forall t > 0$  ( $f''(t) > 0, \forall t > 0$ ) e se  $\forall_{j \neq m} \exists_{i_0} : \lambda_{i_0 j} \neq \lambda_{i_0 m}$  então  $\Delta \xi(\mathbf{z}) < 0$  ( $\Delta \xi(\mathbf{z}) > 0$ ), com  $\Delta$  o operador Laplaciano, isto é  $\xi(\mathbf{z})$  é estritamente superharmónica (subharmónica). Logo pelo princípio do mínimo (máximo)  $\xi(\mathbf{z})$  não pode ter mínimos (máximos) locais no interior de  $\mathcal{S}_{p-1}$ . Aplicando sucessivamente este argumento às restrições de  $\xi(\mathbf{z})$  sobre as fronteiras de  $\mathcal{S}_{p-1}$ , e uma vez que as características do Laplaciano se mantêm, temos que concluir que o mínimo (máximo) global é atingido num vértice, o que corresponde a um vector da base canónica e demonstra **(iii)**.

**(iv)** Sem perda de generalidade vamos considerar apenas o caso  $f''(t) < 0, \forall t > 0$ . Em primeiro lugar é fácil verificar que, sendo  $\xi(\mathbf{z})$  superharmónica, se existir um máximo local no interior de  $\mathcal{S}_{p-1}$  ele terá de ser global. Por outro lado se existir um máximo global no interior terá necessariamente de existir pelo menos um máximo global da restrição de  $\xi(\mathbf{z})$  numa das fronteiras abertas de dimensão 1, (pois se tal não acontecer teremos de concluir que existe pelo menos um ponto de  $\mathcal{S}_{p-1}$  onde a função não é superharmónica). Basta então procurar condições para a existência de máximos globais nas restrições de  $\xi(\mathbf{z})$  às arestas de  $\mathcal{S}_{p-1}$ . Como  $f''(t) < 0$ ,  $f'$  é decrescente e  $\xi'(z_j)$  anula-se para  $0 < z_j < 1$  se e só se  $\xi'_{jm}(0) \times \xi'_{jm}(1) < 0$ , o que conclui a demonstração.  $\square$

### Observações:

1. É possível demonstrar (i) relaxando a condição de diferenciabilidade de  $f$ . No entanto, esta condição não é na prática uma restrição uma vez que as funções que se utilizam são diferenciáveis.
2. Deve notar-se que o facto de os funcionais dos vectores próprios poderem ser, para certas combinações dos valores próprios, diferentes da base canónica, não implica necessariamente um comportamento totalmente descabido em termos amostrais (como acontece nas *componentes principais* usuais quando há valores próprios iguais), pois aqui o que sucede é que se obtém um número finito de soluções cuja média é a solução correcta.

Para simplificar a notação seja  $\mathbf{N}_f$  a matriz simétrica e com diagonal nula tal que  $\mathbf{N}_f(j, m) = (\nu_j - \nu_{jm})(\nu_{mj} - \nu_m)$ . Dados  $f$  e  $\Lambda_i$ ,  $i = 1, \dots, k$ , há problemas de consistência se existir algum  $\mathbf{N}_f(j, m) < 0$ .

## 5.4 Funções de influência parciais e variâncias assintóticas dos estimadores

Nesta subsecção vamos apresentar a dedução das *funções de influência parciais* dos funcionais “projection–pursuit” gerais. Foi novamente o resultado (4.4) que nos permitiu deduzir o teorema que se segue.

### Teorema 5.2 (funções de influência parciais dos funcionais equivalentes aos estimadores PP gerais do modelo CPC)

Sejam  $\mathbf{X}_i$  vectores aleatórios independentes com distribuição elipticamente simétrica  $F_i$ , com parâmetros de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matrizes de dispersão  $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^T$  que satisfazem (2.3). Admita-se que os vectores  $\mathbf{C}_i^{-1} \mathbf{X}_i = \mathbf{Z}_i$  têm distribuição esférica  $G$  para todo  $i = 1, \dots, k$ . Assuma-se que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição de  $Z_{11}$ .

Considere-se  $f$  uma função estritamente crescente contínua e duas vezes diferenciável e seja ainda  $(\varepsilon, y) \rightarrow \sigma((1 - \varepsilon)G_0 + \varepsilon\Delta_y)$  uma função contínua e duas vezes diferenciável em  $(0, y)$ . Então, as *funções de influência parciais* dos funcionais definidos por (5.4), (5.5) e (5.6), são dadas, para todo  $\mathbf{x}$ , por

$$\text{PIF}_i(\mathbf{x}, \lambda_{\sigma, \ell j}, F) = 2 \delta_{\ell i} \lambda_{ij} \text{IF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right), \quad (5.20)$$

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= \tau_i \boldsymbol{\beta}_j^T \mathbf{x} \sum_{s=1}^{j-1} \frac{1}{\nu_{sj} - \nu_s} \sqrt{\lambda_{is}} f'(\lambda_{is}) \text{DIF} \left( \frac{\boldsymbol{\beta}_s^T \mathbf{x}}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) \boldsymbol{\beta}_s + \\ &+ \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \sum_{s=j+1}^p \frac{1}{\nu_j - \nu_{js}} \boldsymbol{\beta}_s^T \mathbf{x} \boldsymbol{\beta}_s, \quad (5.21) \end{aligned}$$

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\nu_{sj} - \nu_s} \tau_i \sqrt{\lambda_{is}} f'(\lambda_{is}) \times \\ &\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) (\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T), \quad (5.22) \end{aligned}$$

onde  $\text{DIF}(y, \sigma, G_0)$  representa a derivada de  $\text{IF}(y, \sigma, G_0)$  em ordem a  $y$ ,  $\nu_{js} = \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{is}$ ,  $\nu_j = \nu_{jj}$  e com  $\nu_{js} \neq \nu_{jj}$  para  $s \neq j$ .

**Dem:** A técnica usada nesta demonstração é semelhante à utilizada no Teorema 4.2.

Seja  $F$  a distribuição definida pela medida produto  $F = F_1 \times \cdots \times F_k$ . Assuma-se que  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e  $\sum_{i=1}^k \tau_i = 1$ .

Represente-se por  $F_{i,\mathbf{x},\varepsilon} = (1-\varepsilon)F_i + \varepsilon \Delta_{\mathbf{x}}$ , com  $F_{\mathbf{x},\varepsilon,i} = F_1 \times \cdots \times F_{i-1} \times F_{i,\mathbf{x},\varepsilon} \times F_{i+1} \times \cdots \times F_k$ , a medida produto com contaminação na  $i$ -ésima população. Considerem-se os seguintes funcionais sob contaminação,  $\boldsymbol{\beta}_{j,\varepsilon,i} = \boldsymbol{\beta}_{\sigma,j}(F_{\mathbf{x},\varepsilon,i})$ ,  $\lambda_{\ell j,\varepsilon,i} = \sigma^2(F_{\ell}[\boldsymbol{\beta}_{j,\varepsilon,i}])$ , para  $\ell \neq i$  e  $\lambda_{ij,\varepsilon,i} = \sigma^2(F_{i,\mathbf{x},\varepsilon}[\boldsymbol{\beta}_{j,\varepsilon,i}])$ .

Considere-se ainda  $\mathbf{V}_{\ell,\varepsilon,i} = \mathbf{V}_{\sigma,\ell}(F_{\mathbf{x},\varepsilon,i}) = \sum_{j=1}^p \lambda_{\ell j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i}^T$ .

• *Funções de influência parciais dos vectores próprios.*

$\boldsymbol{\beta}_{j,\varepsilon,i}$  maximiza  $\varsigma(F_{\mathbf{x},\varepsilon,i}[\mathbf{b}])$  sujeito às condições de ortogonalidade  $\boldsymbol{\beta}_{j,\varepsilon,i}^T \boldsymbol{\beta}_{j,\varepsilon,i} = 1$  e  $\boldsymbol{\beta}_{s,\varepsilon,i}^T \boldsymbol{\beta}_{j,\varepsilon,i} = 0$  para  $s = 1, \dots, j-1$ . Assim,  $\boldsymbol{\beta}_{j,\varepsilon,i}$  maximiza a função Lagrangeana

$$L(\mathbf{b}, \gamma, \boldsymbol{\alpha}) = \tau_i f\{\sigma^2(F_{i,\mathbf{x},\varepsilon}[\mathbf{b}])\} + \sum_{i_0 \neq i} \tau_{i_0} f\{\sigma^2(F_{i_0}[\mathbf{b}])\} - \gamma(\mathbf{b}^T \mathbf{b} - 1) - \sum_{s=1}^{j-1} \alpha_s \mathbf{b}^T \boldsymbol{\beta}_{s,\varepsilon,i}, \quad (5.23)$$

e tem que satisfazer a equação

$$\mathbf{0}_p = \frac{\partial}{\partial \mathbf{b}} L(\mathbf{b}, \gamma, \boldsymbol{\alpha})|_{\mathbf{b}=\boldsymbol{\beta}_{j,\varepsilon,i}} = \psi(\varepsilon) - 2\gamma \boldsymbol{\beta}_{j,\varepsilon,i} - \sum_{s=1}^{j-1} \alpha_s \boldsymbol{\beta}_{s,\varepsilon,i}, \quad (5.24)$$

com

$$\begin{aligned}
\psi(\varepsilon) &= \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{\mathbf{x}, \varepsilon, i}[\mathbf{b}])|_{\mathbf{b}=\boldsymbol{\beta}_{j, \varepsilon, i}} = \\
&= \tau_i \frac{\partial}{\partial \mathbf{b}} f\{\sigma^2(F_{i, \mathbf{x}, \varepsilon}[\mathbf{b}])\}|_{\mathbf{b}=\boldsymbol{\beta}_{j, \varepsilon, i}} + \sum_{i_0 \neq i} \tau_{i_0} \frac{\partial}{\partial \mathbf{b}} f\{\sigma^2(F_{i_0}[\mathbf{b}])\}|_{\mathbf{b}=\boldsymbol{\beta}_{j, \varepsilon, i}} ,
\end{aligned} \tag{5.25}$$

isto é

$$\begin{aligned}
\psi(\varepsilon) &= \tau_i f'\{\sigma^2(F_{i, \mathbf{x}, \varepsilon}[\boldsymbol{\beta}_{j, \varepsilon, i}])\} \frac{\partial}{\partial \mathbf{b}} \{\sigma^2(F_{i, \mathbf{x}, \varepsilon}[\mathbf{b}])\}|_{\mathbf{b}=\boldsymbol{\beta}_{j, \varepsilon, i}} + \\
&+ \sum_{i_0 \neq i} \tau_{i_0} f'\{\sigma^2(\boldsymbol{\beta}_{j, \varepsilon, i}^T \boldsymbol{\Sigma}_{i_0} \boldsymbol{\beta}_{j, \varepsilon, i})\} \frac{\partial}{\partial \mathbf{b}} \{\sigma^2(F_{i_0}[\mathbf{b}])\}|_{\mathbf{b}=\boldsymbol{\beta}_{j, \varepsilon, i_0}} .
\end{aligned} \tag{5.26}$$

De  $\psi(\varepsilon)^T = 2\gamma \boldsymbol{\beta}_{j, \varepsilon, i}^T + \sum_{s=1}^{j-1} \alpha_s \boldsymbol{\beta}_{s, \varepsilon, i}^T$  e da ortogonalidade dos vectores próprios comuns tem-se que

$$\begin{aligned}
\psi(\varepsilon)^T \boldsymbol{\beta}_{j, \varepsilon, i} &= 2\gamma , \\
\psi(\varepsilon)^T \boldsymbol{\beta}_{s, \varepsilon, i} &= \alpha_s , \quad s = 1, \dots, j-1 .
\end{aligned}$$

Portanto, (5.24) pode ser escrita como

$$\psi(\varepsilon) = \sum_{s=1}^j (\psi(\varepsilon)^T \boldsymbol{\beta}_{s, \varepsilon, i}) \boldsymbol{\beta}_{s, \varepsilon, i} . \tag{5.27}$$

Derivando em ordem a  $\varepsilon$  a expressão (5.27), obtém-se

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= \sum_{s=1}^j \left[ (\psi(0)^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma, s}, F)) \boldsymbol{\beta}_s + \left( \boldsymbol{\beta}_s^T \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} \right) \boldsymbol{\beta}_s \right] + \\
&+ \sum_{s=1}^j (\psi(0)^T \boldsymbol{\beta}_s) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma, s}, F) .
\end{aligned} \tag{5.28}$$

Uma vez que  $F_i$  é uma distribuição elipticamente simétrica e como  $\sigma(G_0) = 1$ , usando o facto de  $\sigma^2(F_i[\mathbf{b}]) = \mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}$ , vem

$$\begin{aligned}
\psi(0) &= \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{0, \mathbf{x}, i}[\mathbf{b}])|_{\mathbf{b}=\boldsymbol{\beta}_j} = \sum_{i=1}^k \tau_i \frac{\partial}{\partial \mathbf{b}} f(\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b})|_{\mathbf{b}=\boldsymbol{\beta}_j} = \\
&= 2 \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j f'(\boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j) = \\
&= 2 \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{ij} \boldsymbol{\beta}_j = 2\nu_j \boldsymbol{\beta}_j ,
\end{aligned}$$



o que implica que  $\psi(0)^T \boldsymbol{\beta}_s = 0$  para  $s = 1, \dots, j-1$ . Assim,

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= \sum_{s=1}^j \left[ (\psi(0)^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F)) \boldsymbol{\beta}_s + \left( \boldsymbol{\beta}_s^T \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} \right) \boldsymbol{\beta}_s \right] + \\ &+ (\psi(0)^T \boldsymbol{\beta}_j) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) . \end{aligned} \quad (5.29)$$

Considerando  $\mathbf{P}_{j+1} = \mathbf{I}_p - \sum_{s=1}^j \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T$ , (5.29) pode ser escrita como

$$\mathbf{P}_{j+1} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} = 2\nu_j \sum_{s=1}^j \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s + 2\nu_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) . \quad (5.30)$$

Por outro lado, de (5.26) e uma vez que  $\varsigma(F[\mathbf{b}]) = \sum_{i=1}^k \tau_i f(\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b})$  tem-se pela regra da diferenciação da função composta que

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= \frac{\partial}{\partial \varepsilon} \left( \frac{\partial}{\partial \mathbf{b}} \varsigma(F_{\mathbf{x}, \varepsilon, i}(\mathbf{b})) \right) \Big|_{\mathbf{b}=\boldsymbol{\beta}_{j,i,\varepsilon=0}} + \frac{\partial}{\partial \mathbf{b}} \frac{\partial}{\partial \mathbf{b}^T} \varsigma(F_i(\mathbf{b})|_{\mathbf{b}=\boldsymbol{\beta}_j}) \times \\ &\times \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= 2 \tilde{\boldsymbol{\Sigma}}_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + \\ &+ \tau_i \frac{\partial}{\partial \mathbf{b}} [f' \{ \sigma^2(F_i[\mathbf{b}]) \} \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}])] \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} , \end{aligned} \quad (5.31)$$

onde  $\tilde{\boldsymbol{\Sigma}}_j = \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \boldsymbol{\Sigma}_i = \sum_{s=1}^p \nu_{js} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T$ .

Novamente pela equivariância do funcional de escala, tem-se

$$\begin{aligned} \frac{\partial}{\partial \mathbf{b}} [f' \{ \sigma^2(\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}) \} \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}])] \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} &= \frac{\partial}{\partial \mathbf{b}} [f' \{ \sigma^2(\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}) \}] \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} \times \\ &\times \text{IF}(\boldsymbol{\beta}_j^T \mathbf{x}, \sigma^2, F_i[\boldsymbol{\beta}_j]) + f' \{ \sigma^2(\boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j) \} \frac{\partial}{\partial \mathbf{b}} [\text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}])] \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} , \end{aligned}$$

isto é

$$\begin{aligned} \frac{\partial}{\partial \mathbf{b}} [f' \{ \sigma^2(\mathbf{b}^T \boldsymbol{\Sigma}_i \mathbf{b}) \} \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i[\mathbf{b}])] \Big|_{\mathbf{b}=\boldsymbol{\beta}_j} &= 2\lambda_{ij} f''(\lambda_{ij}) \boldsymbol{\beta}_j \lambda_{ij} \times \\ &\times \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) + f'(\lambda_{ij}) \times \left\{ 2\lambda_{ij} \boldsymbol{\beta}_j \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \right\} + \\ &+ f'(\lambda_{ij}) \times \left\{ \lambda_{ij} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \left( \frac{\mathbf{x}}{\sqrt{\lambda_{ij}}} - \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}} \boldsymbol{\beta}_j \right) \right\} . \end{aligned} \quad (5.32)$$

De (5.31) e (5.32), vem que

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \psi(\varepsilon)|_{\varepsilon=0} &= 2 \tilde{\Sigma}_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + 2\tau_i \{f''(\lambda_{ij}) \lambda_{ij} + f'(\lambda_{ij})\} \lambda_{ij} \boldsymbol{\beta}_j \times \\
&\times \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) + \tau_i f'(\lambda_{ij}) \sqrt{\lambda_{ij}} \times \\
&\times \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) (\mathbf{x} - \boldsymbol{\beta}_j^T \mathbf{x} \boldsymbol{\beta}_j) .
\end{aligned} \tag{5.33}$$

Assim, em consequência de (5.33), (5.30) pode ser escrita como

$$\begin{aligned}
2\nu_j \sum_{s=1}^j \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s + 2\nu_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= \\
&= 2 \mathbf{P}_{j+1} \tilde{\Sigma}_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) + \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \times \\
&\times \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \mathbf{P}_{j+1} \mathbf{x} .
\end{aligned} \tag{5.34}$$

De  $\boldsymbol{\beta}_{j,\varepsilon}^T \boldsymbol{\beta}_{j,\varepsilon} = 1$  resulta que  $\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T \boldsymbol{\beta}_j = 0$ , e portanto (5.34) é equivalente a

$$\begin{aligned}
2 \left( \mathbf{P}_{j+1} \tilde{\Sigma}_j - \nu_{jp} \right) \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= 2\nu_j \sum_{s=1}^{j-1} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s - \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \times \\
&\times \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0 \right) \mathbf{P}_{j+1} \mathbf{x} .
\end{aligned} \tag{5.35}$$

Por outro lado, tem-se que

$$\begin{aligned}
\mathbf{P}_{j+1} \tilde{\Sigma}_j &= \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \boldsymbol{\Sigma}_i - \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \boldsymbol{\Sigma}_i \sum_{s=1}^j \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\
&= \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \boldsymbol{\Sigma}_i - \sum_{i=1}^k \sum_{s=1}^j \tau_i f'(\lambda_{ij}) \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\
&= \sum_{i=1}^k \sum_{s=1}^p \tau_i f'(\lambda_{ij}) \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T - \sum_{i=1}^k \sum_{s=1}^j \tau_i f'(\lambda_{ij}) \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\
&= \sum_{i=1}^k \sum_{s=j+1}^p \tau_i f'(\lambda_{ij}) \lambda_{is} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T = \\
&= \sum_{s=j+1}^p \left( \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{is} \right) \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T .
\end{aligned}$$

Uma vez que  $\sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{is} = \nu_{js}$ , vem

$$\mathbf{P}_{j+1} \tilde{\Sigma}_j = \sum_{s=j+1}^p \nu_{js} \boldsymbol{\beta}_s \boldsymbol{\beta}_s^T . \tag{5.36}$$

Resulta então que a matriz  $\mathbf{P}_{j+1}\tilde{\Sigma}_j - \nu_j\mathbf{I}_p = \sum_{s=j+1}^p \nu_{js}\boldsymbol{\beta}_s\boldsymbol{\beta}_s^T - \nu_j\mathbf{I}_p$  tem característica completa, pois  $\nu_{js} \neq \nu_{jj}$  para  $s \neq j$ , com inversa

$$\left(\mathbf{P}_{j+1}\tilde{\Sigma}_j - \nu_j\mathbf{I}_p\right)^{-1} = \sum_{s=j+1}^p \frac{1}{\nu_{js} - \nu_j} \boldsymbol{\beta}_s\boldsymbol{\beta}_s^T - \sum_{s=1}^j \frac{1}{\nu_j} \boldsymbol{\beta}_s\boldsymbol{\beta}_s^T,$$

e portanto

$$\left(\mathbf{P}_{j+1}\tilde{\Sigma}_j - \nu_j\mathbf{I}_p\right)^{-1} \boldsymbol{\beta}_s = -\frac{1}{\nu_j} \boldsymbol{\beta}_s,$$

para  $s = 1, \dots, j-1$ . Assim, tem-se

$$\left(\mathbf{P}_{j+1}\tilde{\Sigma}_j - \nu_j\mathbf{I}_p\right)^{-1} \mathbf{P}_{j+1} = \sum_{s=j+1}^p \frac{1}{\nu_{js} - \nu_j} \boldsymbol{\beta}_s\boldsymbol{\beta}_s^T.$$

A aplicação do resultado anterior em (5.35) conduz a

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= -\sum_{s=1}^{j-1} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F) \boldsymbol{\beta}_s - \frac{1}{2} \sum_{s=j+1}^p \frac{1}{\nu_{js} - \nu_j} \boldsymbol{\beta}_s \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \times \\ &\times \text{DIF}\left(\frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right) \boldsymbol{\beta}_s^T \mathbf{x}. \end{aligned} \quad (5.37)$$

De (5.37), para  $s = j+1, \dots, p$  vem que

$$\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T \boldsymbol{\beta}_s = \frac{1}{2(\nu_j - \nu_{js})} \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \text{DIF}\left(\frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right) \boldsymbol{\beta}_s^T \mathbf{x}.$$

Como consequência,

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= -\tau_i \boldsymbol{\beta}_j^T \mathbf{x} \sum_{s=1}^{j-1} \frac{1}{2(\nu_s - \nu_{sj})} \sqrt{\lambda_{is}} f'(\lambda_{is}) \times \\ &\times \text{DIF}\left(\frac{\boldsymbol{\beta}_s^T \mathbf{x}}{\sqrt{\lambda_{is}}}, \sigma^2, G_0\right) \boldsymbol{\beta}_s - \frac{1}{2} \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \times \\ &\times \text{DIF}\left(\frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right) \sum_{s=j+1}^p \frac{1}{\nu_{js} - \nu_j} \boldsymbol{\beta}_s^T \mathbf{x} \boldsymbol{\beta}_s. \end{aligned} \quad (5.38)$$

Finalmente, usando o facto de  $\text{IF}(y, \sigma^2, G_0) = 2\text{IF}(y, \sigma, G_0)$ , obtém-se o resultado (5.21).

- *Funções de influência parciais dos valores próprios.*

Usando o facto de  $\lambda_{ij,\varepsilon,i} = \sigma^2 (F_{i,\mathbf{x},\varepsilon} [\boldsymbol{\beta}_{j,\varepsilon,i}])$  para  $\ell \neq i$ , e de  $\lambda_{\ell j,\varepsilon,i} = \sigma^2 (F_\ell [\boldsymbol{\beta}_{j,\varepsilon,i}]) = \boldsymbol{\beta}_{j,\varepsilon,i}^T \boldsymbol{\Sigma}_\ell \boldsymbol{\beta}_{j,\varepsilon,i}$ , com a regra da diferenciação da função composta, resulta

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \lambda_{\sigma,ij}, F) &= \frac{\partial}{\partial \varepsilon} \sigma^2 (F_{i,\varepsilon,\mathbf{x}} [\boldsymbol{\beta}_j]) \big|_{\varepsilon=0} + \left[ \frac{\partial}{\partial \mathbf{b}} \sigma^2 (F_i [\mathbf{b}]) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} \right]^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i(\mathbf{b})) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} + (2\boldsymbol{\Sigma}_i \boldsymbol{\beta}_j)^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= \text{IF}(\mathbf{b}^T \mathbf{x}, \sigma^2, F_i(\mathbf{b})) \big|_{\mathbf{b}=\boldsymbol{\beta}_j} + 2\lambda_{ij} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = \\ &= 2\lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right), \end{aligned}$$

para  $i = 1, \dots, k$ ,  $j = 1, \dots, p$ .

Então, vem que  $\text{PIF}_i(\mathbf{x}, \lambda_{\sigma,\ell j}, F) = 2\lambda_{\ell j} \boldsymbol{\beta}_j^T \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = 0$ , para  $i = 1, \dots, k$  e  $\ell \neq i$ , o que conclui a prova de (5.20).

• *Funções de influência parciais das matrizes de dispersão.*

Como  $\mathbf{V}_{\ell,\varepsilon,i} = \mathbf{V}_{\sigma,\ell}(F_{\varepsilon,\mathbf{x},i}) = \sum_{j=1}^p \lambda_{\ell j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i}^T$ , e usando as expressões das *funções de influência parciais* para os vectores e valores próprios, vem que

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) &= \sum_{j=1}^p \text{PIF}_i(\mathbf{x}, \lambda_{\sigma,\ell j}, F) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \sum_{j=1}^p \lambda_{\ell j} \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=1}^p \lambda_{\ell j} \boldsymbol{\beta}_j \text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)^T, \end{aligned}$$

portanto

$$\begin{aligned} \text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) &= 2 \delta_{\ell i} \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \lambda_{\ell j} \frac{1}{\nu_{sj} - \nu_s} \tau_i \sqrt{\lambda_{is}} f'(\lambda_{is}) \times \\ &\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) (\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T) + \\ &+ \sum_{j=2}^p \sum_{s=1}^{j-1} \lambda_{\ell s} \frac{1}{\nu_s - \nu_{sj}} \tau_i \sqrt{\lambda_{is}} f'(\lambda_{is}) \times \\ &\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) (\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T), \end{aligned}$$

resultando finalmente

$$\begin{aligned}
\text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma, \ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\
&+ \sum_{j=2}^p \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\nu_{sj} - \nu_s} \tau_i \sqrt{\lambda_{is}} f'(\lambda_{is}) \times \\
&\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) [\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T] ,
\end{aligned}$$

o que conclui a prova.  $\square$

É fácil constatar, através da observação de (5.20), que para os valores próprios as *funções de influência parciais* são limitadas caso se considere um funcional de escala com *função de influência* limitada. Esta situação é exactamente a mesma que a observada em (4.49) com a aplicação da função identidade, dado que as *funções de influência parciais* dos valores próprios não dependem da função  $f$  considerada. Por outro lado, no que se refere às *funções de influência parciais* dos vectores próprios estas mantêm-se ilimitadas para qualquer  $f$  devido ao efeito do termo  $\mathbf{x}^T \boldsymbol{\beta}_j$  (que não pode ser compensado por  $\text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right)$  uma vez que este actua numa direcção ortogonal).

No caso particular de  $f(t) = \ln(t)$  resulta

$$\begin{aligned}
\text{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma, j}, F) &= \tau_i \boldsymbol{\beta}_j^T \mathbf{x} \sum_{s=1}^{j-1} \frac{1}{\nu_{sj} - 1} \frac{1}{\sqrt{\lambda_{is}}} \text{DIF} \left( \frac{\boldsymbol{\beta}_s^T \mathbf{x}}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) \boldsymbol{\beta}_s + \\
&+ \frac{\tau_i}{\sqrt{\lambda_{ij}}} \text{DIF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \sum_{s=j+1}^p \frac{1}{1 - \nu_{js}} \boldsymbol{\beta}_s^T \mathbf{x} \boldsymbol{\beta}_s , \quad (5.39)
\end{aligned}$$

$$\begin{aligned}
\text{PIF}_i(\mathbf{x}, \mathbf{V}_{\sigma, \ell}, F) &= \delta_{\ell i} 2 \sum_{j=1}^p \lambda_{ij} \text{IF} \left( \frac{\boldsymbol{\beta}_j^T \mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j^T + \\
&+ \sum_{j=2}^p \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\nu_{sj} - 1} \frac{\tau_i}{\sqrt{\lambda_{is}}} \times \\
&\times \text{DIF} \left( \frac{\mathbf{x}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{x}^T \boldsymbol{\beta}_j) (\boldsymbol{\beta}_s \boldsymbol{\beta}_j^T + \boldsymbol{\beta}_j \boldsymbol{\beta}_s^T) . \quad (5.40)
\end{aligned}$$

Confrontando (5.39) com (4.50), observa-se que a introdução da função logaritmo originou a pré-multiplicação do termo  $\mathbf{x}^T \boldsymbol{\beta}_j$  pelo inverso da raiz quadrada da variância das *componentes principais comuns*. À medida que a variabilidade dos eixos comuns aumenta o efeito do termo  $\mathbf{x}^T \boldsymbol{\beta}_j$  vai diminuindo, conduzindo a uma menor influência de contaminações pontuais desfavoráveis, contrariamente ao que se verifica com os estimadores PP usuais construídos com a função identidade.

As variâncias assíntóticas dos estimadores PP gerais dos valores e vectores próprios podem ser obtidas heurísticamente através de (4.6). Estes resultados são apresentados no corolário que se segue.

**Corolário 5.1 (variâncias assíntóticas dos estimadores PP gerais do modelo CPC)**

Sejam  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, \dots, k$ , vectores aleatórios independentes de  $k$  amostras independentes com distribuição  $F_i$ . Seja  $\boldsymbol{\mu}_i = \mathbf{0}_p$  o parâmetro de localização e  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$  a matriz de dispersão que satisfaz o modelo (2.3) com  $\boldsymbol{\beta} = \mathbf{I}_p$ . Assuma-se também que  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$  e  $\sum_{i=1}^k \tau_i = 1$ .

Considere-se que  $\boldsymbol{\Lambda}_i^{-\frac{1}{2}} \mathbf{X}_{i1} = \mathbf{Z}_i$  tem a mesma distribuição esférica  $G$  para todo o  $i = 1, \dots, k$ . Assuma-se ainda que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição de  $Z_{11}$ .

Represente-se por  $s(\cdot)$  o estimador de escala univariado relacionado com  $\sigma(\cdot)$  e seja  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i})$  para  $i = 1, \dots, k$ . Os estimadores dos eixos principais comuns definem-se com a resolução iterativa de (5.3), enquanto que os estimadores dos valores próprios e da matriz de dispersão da  $i$ -ésima população são calculados por

$$\hat{\lambda}_{ij} = s^2(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_j), \quad \text{para } j = 1, \dots, p, \quad \mathbf{V}_i = \sum_{j=1}^p \hat{\lambda}_{ij} \hat{\boldsymbol{\beta}}_j \hat{\boldsymbol{\beta}}_j^T. \quad (5.41)$$

Então, caso  $f$  seja duas vezes diferenciável e a função  $(\varepsilon, y) \rightarrow \sigma((1 - \varepsilon)G_0 + \varepsilon\Delta_y)$  seja contínua e também duas vezes diferenciável em  $(0, y)$ , as variâncias assíntóticas dos estimadores PP solução de (5.3) e (5.41) são dadas por

$$\text{ASVAR}(\hat{\lambda}_{\ell j}) = 4\lambda_{\ell j}^2 \frac{1}{\tau_\ell} \text{ASVAR}(s, G_0), \quad (5.42)$$

$$\begin{aligned} \text{ASVAR}(\hat{\beta}_{jm}) &= \sum_{i=1}^k \tau_i \lambda_{ij} \lambda_{im} \left\{ \frac{\delta_{m>j} \{f'(\lambda_{ij})\}^2}{(\nu_j - \nu_{jm})^2} + \frac{\delta_{m<j} \{f'(\lambda_{im})\}^2}{(\nu_{mj} - \nu_m)^2} \right\} \times \\ &\times E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2, \quad m \neq j, \end{aligned} \quad (5.43)$$

$$\text{ASVAR}(\hat{\beta}_{jm}) = 0, \quad m = j. \quad (5.44)$$

Em particular, no caso de  $G = N(\mathbf{0}_p, \mathbf{I}_p)$ , tem-se que

$$\begin{aligned} \text{ASVAR}(\hat{\lambda}_{\ell j}) &= 4\lambda_{\ell j}^2 \frac{1}{\tau_{\ell}} \text{ASVAR}(s, \Phi) , \\ \text{ASVAR}(\hat{\beta}_{jm}) &= \sum_{i=1}^k \tau_i \lambda_{ij} \lambda_{im} \left\{ \delta_{m>j} \frac{\{f'(\lambda_{ij})\}^2}{(\nu_j - \nu_{jm})^2} + \delta_{m<j} \frac{\{f'(\lambda_{im})\}^2}{(\nu_{mj} - \nu_m)^2} \right\} \times \\ &\quad \times E_{\Phi} [\text{DIF}(Y, \sigma, \Phi)]^2 , \quad \text{para } m \neq j , \\ \text{ASCOV}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) &= 0 , \quad \text{para } m \neq j , \quad m \neq r , \quad r \neq j , \end{aligned} \quad (5.45)$$

onde  $\nu_{js} = \sum_{i=1}^k \tau_i f'(\lambda_{ij}) \lambda_{is}$ ,  $\nu_j = \nu_{jj}$  e  $\nu_{js} \neq \nu_{jj}$  para  $s \neq j$ .

**Dem:** De (4.6) vem que

$$\begin{aligned} \text{ASVAR}(\hat{\lambda}_{\ell j}) &= \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} [\text{PIF}_i(\mathbf{X}, \lambda_{\sigma, \ell j}, F)^2] , \\ \text{ASVAR}(\hat{\beta}_{jm}) &= \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} [\text{PIF}_i(\mathbf{X}, \beta_{\sigma, jm}, F)^2] . \end{aligned}$$

Usando (5.20) em conjunto com o facto de  $\frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{\ell j}}} \sim G_0$  quando  $\mathbf{X} \sim F_i$ , obtém-se

$$\begin{aligned} \text{ASVAR}(\hat{\lambda}_{\ell j}) &= \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} \left[ \left\{ \delta_{\ell i} 2\lambda_{ij} \text{IF} \left( \frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \right\}^2 \right] = \\ &= 4\lambda_{\ell j}^2 \frac{1}{\tau_{\ell}} \text{ASVAR}(s, G_0) . \end{aligned}$$

A partir de (5.21) e considerando que  $\boldsymbol{\beta} = \mathbf{I}_p$  resulta

$$\begin{aligned} \text{ASVAR}(\hat{\beta}_{jm}) &= \sum_{i=1}^k \frac{1}{\tau_i} E_{F_i} \left\{ \left[ \tau_i \sqrt{\lambda_{ij}} f'(\lambda_{ij}) \text{DIF} \left( \frac{\mathbf{X}^T \boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0 \right) \right]^2 \times \right. \\ &\quad \times \left[ \sum_{s=j+1}^p \frac{1}{\nu_j - \nu_{js}} \beta_{sm} (\mathbf{X}^T \boldsymbol{\beta}_s) \right]^2 + \\ &\quad \left. + \left[ \tau_i \sum_{s=1}^{j-1} \frac{1}{\nu_{sj} - \nu_s} \beta_{sm} \sqrt{\lambda_{is}} f'(\lambda_{is}) \text{DIF} \left( \frac{\mathbf{X}^T \boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0 \right) (\mathbf{X}^T \boldsymbol{\beta}_j) \right]^2 \right\} . \end{aligned}$$

Assim

$$\begin{aligned} \text{ASVAR}(\hat{\beta}_{jm}) &= \delta_{m>j} \sum_{i=1}^k \tau_i \frac{\{f'(\lambda_{ij})\}^2 \lambda_{ij} \lambda_{im}}{(\nu_j - \nu_{jm})^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 + \\ &\quad + \delta_{m<j} \sum_{i=1}^k \tau_i \frac{\{f'(\lambda_{im})\}^2 \lambda_{ij} \lambda_{im}}{(\nu_{mj} - \nu_m)^2} E_G \{ \text{DIF}(Z_{1m}, \sigma, G_0) Z_{1j} \}^2 , \end{aligned}$$

resultando finalmente

$$\begin{aligned} \text{ASVAR} \left( \widehat{\beta}_{jm} \right) &= \sum_{i=1}^k \tau_i \lambda_{ij} \lambda_{im} \left\{ \delta_{m>j} \frac{\{f'(\lambda_{ij})\}^2}{(\nu_j - \nu_{jm})^2} + \delta_{m<j} \frac{\{f'(\lambda_{im})\}^2}{(\nu_{mj} - \nu_m)^2} \right\} \times \\ &\times E_G \{ \text{DIF} (Z_{1j}, \sigma, G_0) Z_{1m} \}^2, \end{aligned}$$

onde  $\delta_{m>j} = 0$  caso  $m \leq j$  e  $\delta_{m>j} = 1$  se  $m > j$ .

Por outro lado, no caso particular de  $G = N(\mathbf{0}_p, \mathbf{I}_p)$ , e uma vez que  $Z_{1j}$ ,  $Z_{1m}$  e  $Z_{1r}$  são variáveis aleatórias independentes se  $j \neq r$ ,  $j \neq m$  e  $m \neq r$  e  $E_\Phi [\text{DIF} (Y, \sigma, \Phi)] = 0$ , resulta então

$$\begin{aligned} \text{ASCOV} \left( \widehat{\beta}_{jm}, \widehat{\beta}_{jr} \right) &= \delta_{m>j} \delta_{r>j} \sum_{i=1}^k \tau_i \frac{\{f'(\lambda_{ij})\}^2 \lambda_{ij} \sqrt{\lambda_{im}} \sqrt{\lambda_{ir}}}{(\nu_j - \nu_{jm})(\nu_j - \nu_{jr})} \times \\ &\times E_G \left[ \{ \text{DIF} (Z_{1j}, \sigma, G_0) \}^2 \{ Z_{1m} z_{1r} \} \right] + \\ &+ \delta_{m<j} \delta_{r<j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \sqrt{\lambda_{im}} \sqrt{\lambda_{ir}} f'(\lambda_{ir}) f'(\lambda_{im})}{(\nu_{mj} - \nu_m)(\nu_{rj} - \nu_r)} \times \\ &\times E_G \left\{ \text{DIF} (Z_{1m}, \sigma, G_0) \text{DIF} (Z_{1r}, \sigma, G_0) (Z_{1j})^2 \right\} + \\ &+ \delta_{m>j} \delta_{r<j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \sqrt{\lambda_{im}} \sqrt{\lambda_{ir}} f'(\lambda_{ij}) f'(\lambda_{ir})}{(\nu_j - \nu_{jm})(\nu_{rj} - \nu_r)} \times \\ &\times E_G \left\{ \text{DIF} (Z_{1j}, \sigma, G_0) \text{DIF} (Z_{1r}, \sigma, G_0) Z_{1m} Z_{1j} \right\} + \\ &+ \delta_{m<j} \delta_{r>j} \sum_{i=1}^k \tau_i \frac{\lambda_{ij} \sqrt{\lambda_{ir}} \sqrt{\lambda_{im}} f'(\lambda_{ij}) f'(\lambda_{im})}{(\nu_{mj} - \nu_m)(\nu_j - \nu_{jr})} \times \\ &\times E_G \left\{ \text{DIF} (Z_{1j}, \sigma, G_0) \text{DIF} (Z_{1m}, \sigma, G_0) Z_{1r} Z_{1j} \right\}, \end{aligned}$$

ficando  $\text{ASCOV} \left( \widehat{\beta}_{jm}, \widehat{\beta}_{jr} \right) = 0$ , o que conclui a demonstração.  $\square$

Em particular, quando  $f(t) = \ln(t)$  é fácil concluir que a variância assintótica dos estimadores dos vectores próprios comuns fica

$$\begin{aligned} \text{ASVAR} \left( \widehat{\beta}_{jm} \right) &= \sum_{i=1}^k \tau_i \left\{ \delta_{m<j} \frac{\lambda_{im}}{\lambda_{ij}} \frac{1}{(1 - \nu_{jm})^2} + \delta_{m>j} \frac{\lambda_{ij}}{\lambda_{im}} \frac{1}{(1 - \nu_{mj})^2} \right\} \times \\ &\times E_G \{ \text{DIF} (Z_{1j}, \sigma, G_0) Z_{1m} \}^2. \end{aligned} \quad (5.46)$$

É ainda interessante referir que, caso todas as populações tenham a mesma matriz de dispersão a variância assintótica dos estimadores PP dos vectores próprios é



independente da função  $f$ . Nesta situação, para  $m > j$  vem que

$$\begin{aligned} \text{ASVAR}(\hat{\beta}_{jm}) &= \frac{\sum_{i=1}^k \tau_i \lambda_j \lambda_m f'(\lambda_j)^2}{\left( \sum_{i=1}^k \tau_i f'(\lambda_j) (\lambda_j - \lambda_m) \right)^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 = \\ &= \frac{\lambda_j \lambda_m}{(\lambda_j - \lambda_m)^2} E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2, \end{aligned} \quad (5.47)$$

obtendo-se a mesma expressão para  $m < j$ . Este resultado coincide com o obtido por Croux e Ruiz-Gazen (2000) para o caso das *componentes principais*, como não podia deixar de ser. Refira-se ainda que para este caso particular os estimadores PP gerais coincidem com os estimadores PP usuais.

## 5.5 A optimalidade da função logaritmo

O nosso estudo inciduiu, com maior profundidade, sobre os estimadores “projection-pursuit” construídos com a aplicação da função logaritmo. Como se viu anteriormente, estes estimadores surgiram na sequência das equações de verosimilhança de Flury (1984) e portanto aproximam-se da abordagem deste autor. Foi possível provar que com estes estimadores PP, construídos com o logaritmo da variabilidade combinada das  $k$  variáveis projectadas, se atinge a variância assintótica mínima sob a validade do modelo proporcional. Este resultado é mais formalmente apresentado na proposição que se segue. Relembre-se que a estrutura de dispersão inerente ao modelo proporcional, apresentada em (2.2), é mais exigente que a do modelo CPC porque impõe que as  $k$  matrizes de dispersão se relacionem apenas através de uma constante de proporcionalidade.

### Proposição 5.1 (optimalidade da função logaritmo para $k$ populações proporcionais)

Sejam  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, \dots, k$ , vectores aleatórios independentes de  $k$  amostras aleatórias independentes com distribuição  $F_i$ , com parâmetro de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matriz de dispersão  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i = \rho_i \boldsymbol{\Lambda}_1$ , com  $\boldsymbol{\Sigma}_i$  satisfazendo o modelo proporcional (2.2) com  $\boldsymbol{\beta} = \mathbf{I}_p$ ,  $\rho_1 = 1$ . Considere-se  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$ ,  $\sum_{i=1}^k \tau_i = 1$  e  $N = \sum_{i=1}^k n_i$ . Seja  $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$  e assumase que  $\lambda_1 > \dots > \lambda_p$ .

Considere-se ainda que  $\boldsymbol{\Lambda}_i^{-\frac{1}{2}} \mathbf{X}_{i1} = \mathbf{Z}_i$  tem a mesma distribuição esférica  $G$  para

todo o  $i = 1, \dots, k$ . Assuma-se que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição da variável aleatória  $Z_{11}$ .

Denote-se por  $s(\cdot)$  um estimador de escala univariado relacionado com o funcional de escala  $\sigma(\cdot)$  para o qual a função  $(\varepsilon, y) \rightarrow \sigma((1 - \varepsilon)G_0 + \varepsilon\Delta_y)$  é contínua e duas vezes diferenciável em  $(0, y)$ . Então, os estimadores definidos por (5.3), para  $f$  crescente e duas vezes diferenciável, atingem o valor mínimo da variância assintótica dada por (5.43) quando  $f(t) = \ln(t)$ .

**Dem:** Por uma questão de simplicidade, considera-se apenas a situação de  $m > j$ . A demonstração para o caso de  $m < j$  é análoga.

Para o modelo proporcional o facto de  $\lambda_{im} = \rho_i \lambda_m$  leva à igualdade

$$(\nu_{jm} - \nu_j) = \sum_{i=1}^k \tau_i \rho_i f'(\rho_i \lambda_j) (\lambda_m - \lambda_j) . \quad (5.48)$$

Aplicando (5.48) em (5.43), resulta

$$\begin{aligned} \text{ASVAR}(\hat{\beta}_{jm}) &= \frac{\lambda_j \lambda_m}{(\lambda_m - \lambda_j)^2} \frac{\sum_{i=1}^k \tau_i \rho_i^2 [f'(\rho_i \lambda_j)]^2}{\left\{ \sum_{i=1}^k \tau_i \rho_i f'(\rho_i \lambda_j) \right\}^2} \times \\ &\times E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2 . \end{aligned} \quad (5.49)$$

Portanto, de (5.49), é evidente que o mínimo da variância assintótica é atingido para uma função  $f$  estritamente crescente que minimiza

$$\begin{aligned} V(f) &= \frac{\sum_{i=1}^k \tau_i \rho_i^2 \{f'(\rho_i \lambda_j)\}^2}{\left\{ \sum_{i=1}^k \tau_i \rho_i f'(\rho_i \lambda_j) \right\}^2} = \frac{\sum_{i=1}^k \left\{ \tau_i^{\frac{1}{2}} \rho_i f'(\rho_i \lambda_j) \right\}^2}{\left\{ \sum_{i=1}^k \tau_i^{\frac{1}{2}} \tau_i^{\frac{1}{2}} \rho_i f'(\rho_i \lambda_j) \right\}^2} = \\ &= \frac{\|\mathbf{y}_j\|^2}{(\mathbf{y}_j^T \boldsymbol{\xi})^2} , \end{aligned} \quad (5.50)$$

onde  $\mathbf{y}_j = \left( \tau_1^{\frac{1}{2}} \rho_1 f'(\rho_1 \lambda_j), \dots, \tau_k^{\frac{1}{2}} \rho_k f'(\rho_k \lambda_j) \right)^T$  e  $\boldsymbol{\xi} = \left( \tau_1^{\frac{1}{2}}, \dots, \tau_k^{\frac{1}{2}} \right)^T$ . Da desigualdade de Cauchy-Schwartz resulta que  $(\mathbf{y}_j^T \boldsymbol{\xi})^2 \leq \|\mathbf{y}_j\|^2 \|\boldsymbol{\xi}\|^2$  e assim  $V(f) \geq \|\boldsymbol{\xi}\|^{-2} = 1$ .

Por outro lado,

$$\begin{aligned} V(\ln) &= \frac{\sum_{i=1}^k \tau_i \rho_i^2 (\rho_i \lambda_j)^{-2}}{\left\{ \sum_{i=1}^k \tau_i \rho_i (\rho_i \lambda_j)^{-1} \right\}^2} = \frac{(\lambda_j)^{-2} \sum_{i=1}^k \tau_i}{(\lambda_j)^{-2} \left( \sum_{i=1}^k \tau_i \right)^2} = 1 , \end{aligned} \quad (5.51)$$

o que conclui a prova.  $\square$

A investigação prosseguiu no sentido de se tentar aprofundar as propriedades destes estimadores PP particulares, sob a validade do modelo (2.2). Centrando o estudo sobre a variância assintótica, foi ainda possível demonstrar que, restringindo a classe de funções onde  $f$  pode ser escolhida, a única função que minimiza a variância assintótica é  $f(t) = \ln(t)$ . Este resultado apresenta-se na proposição que se segue.

**Proposição 5.2 (unicidade da optimalidade da função logaritmo para populações proporcionais)**

Sejam  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ ,  $i = 1, \dots, k$ , vectores aleatórios independentes de  $k$  amostras aleatórias independentes com distribuição  $F_i$ , com parâmetro de localização  $\boldsymbol{\mu}_i = \mathbf{0}_p$  e matriz de dispersão  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i = \rho_i \boldsymbol{\Lambda}_1$ , com  $\boldsymbol{\Sigma}_i$  satisfazendo o modelo proporcional (2.2) com  $\boldsymbol{\beta} = \mathbf{I}_p$ ,  $\rho_1 = 1$ . Considere-se  $n_i = \tau_i N$ , com  $0 < \tau_i < 1$ ,  $\sum_{i=1}^k \tau_i = 1$  e  $N = \sum_{i=1}^k n_i$ . Seja  $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$  e assumase que  $\lambda_1 > \dots > \lambda_p$ .

Considere-se ainda que  $\boldsymbol{\Lambda}_i^{-\frac{1}{2}} \mathbf{X}_{i1} = \mathbf{Z}_i$  tem a mesma distribuição esférica  $G$  para todo o  $i = 1, \dots, k$ . Assuma-se que  $\sigma(G_0) = 1$ , onde  $G_0$  representa a distribuição da variável aleatória  $Z_{11}$ .

Denote-se por  $s(\cdot)$  um estimador de escala univariado relacionado com o funcional de escala  $\sigma(\cdot)$  e para o qual a função  $(\varepsilon, y) \rightarrow \sigma((1 - \varepsilon)G_0 + \varepsilon\Delta_y)$  é contínua e duas vezes diferenciável em  $(0, y)$ . Então, dentro da classe de funções  $\mathcal{F} = \{f(t) : f'(t) = t^{\alpha-1}, \alpha \geq 0\}$  a única função para a qual os estimadores definidos por (5.3) atingem o valor mínimo da variância assintótica, dada por (5.43), é  $f(t) = \ln(t)$ .

**Dem:** Novamente por uma questão de simplicidade considera-se só a situação de  $m < j$ .

Considerando (5.43) para a classe de funções  $\mathcal{F} = \{f(t) : f'(t) = t^{\alpha-1}, \alpha \geq 0\}$ , resulta

$$\text{ASVAR}(\hat{\beta}_{jm}) = \frac{\sum_{i=1}^k \tau_i \lambda_{ij} \lambda_{im} \lambda_{ij}^{2\alpha-2}}{\left\{ \sum_{i=1}^k \tau_i \lambda_{ij}^{\alpha-1} (\lambda_{ij} - \lambda_{im}) \right\}^2} \times E_G \{ \text{DIF}(Z_{1j}, \sigma, G_0) Z_{1m} \}^2. \quad (5.52)$$

Assim, o processo de pesquisa do mínimo da variância assintótica passa pela

procura do valor de  $\alpha$  que minimiza

$$V(\alpha) = \frac{\sum_{i=1}^k \tau_i a_i \lambda_{ij}^{2\alpha}}{\left\{ \sum_{i=1}^k \tau_i (1 - a_i) \lambda_{ij}^\alpha \right\}^2}, \quad (5.53)$$

com  $a_i = \frac{\lambda_{im}}{\lambda_{ij}}$ ,  $i = 1, \dots, k$  (notar que existe pelo menos um  $a_i \neq 1$ ). Derivando esta expressão em ordem a  $\alpha$ , notando que os termos com  $i = \ell$  se anulam e agrupando os termos com índices simétricos, obtém-se

$$\begin{aligned} V'(\alpha) &= \frac{\sum_{i=1}^k \sum_{\ell=1}^k \left\{ \tau_i \tau_\ell a_i (1 - a_\ell) \log(\lambda_{ij}^2) \lambda_{ij}^{2\alpha} \lambda_{\ell j}^\alpha - \tau_i \tau_\ell a_\ell (1 - a_i) \log(\lambda_{ij}^2) \lambda_{\ell j}^{2\alpha} \lambda_{ij}^\alpha \right\}}{\left\{ \sum_{i=1}^k \tau_i (1 - a_i) \lambda_{ij}^\alpha \right\}^3} = \\ &= \frac{\sum_{i=1}^k \sum_{\ell=1}^{i-1} \tau_i \tau_\ell \lambda_{ij}^\alpha \lambda_{\ell j}^\alpha [\log(\lambda_{ij}^2) - \log(\lambda_{\ell j}^2)] [a_i (1 - a_\ell) \lambda_{ij}^\alpha - a_\ell (1 - a_i) \lambda_{\ell j}^\alpha]}{\left\{ \sum_{i=1}^k \tau_i (1 - a_i) \lambda_{ij}^\alpha \right\}^3}. \end{aligned} \quad (5.54)$$

No caso do modelo proporcional, que se está a considerar, tem-se  $a_1 = \dots = a_k = \frac{\lambda_m}{\lambda_j}$ , pelo que a expressão anterior se pode escrever como

$$V'(\alpha) = \frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2} \frac{\sum_{i=1}^k \sum_{\ell=1}^{i-1} \tau_i \tau_\ell \lambda_{ij}^\alpha \lambda_{\ell j}^\alpha (\lambda_{ij}^\alpha - \lambda_{\ell j}^\alpha) [\log(\lambda_{ij}^2) - \log(\lambda_{\ell j}^2)]}{\left\{ \sum_{i=1}^k \tau_i \lambda_{ij}^\alpha \right\}^3}. \quad (5.55)$$

Como todas as parcelas de (5.55) são maiores ou iguais a zero conclui-se que  $V'(\alpha) = 0$  se e só se todas as parcelas se anularem, ou seja (como  $\tau_i \neq 0 \ \forall i$  e  $\lambda_{ij} \neq 0 \ \forall i, j$ ) se e só se

$$(\lambda_{ij}^\alpha - \lambda_{\ell j}^\alpha) [\log(\lambda_{ij}^2) - \log(\lambda_{\ell j}^2)] = 0, \ \forall i, \ell, \quad (5.56)$$

o que é equivalente a  $(\lambda_{ij} - \lambda_{\ell j}) = 0, \ \forall i, \ell$ , ou  $\alpha = 0$ . A primeira condição é impossível porque  $\exists_{i_0} : \rho_{i_0} \neq 1$ , conclui-se então que  $V'(0) = 0$  e que  $V'(\alpha) > 0, \ \forall \alpha > 0$ , o que é equivalente a dizer que  $V(\alpha)$  atinge o mínimo apenas para  $\alpha = 0$  como se queria demonstrar.  $\square$

### Observações:

1. Como se viu as *funções de influência parciais* dos funcionais valores próprios não dependem da função  $f$ . Assim, sob o modelo proporcional, a aplicação

da função  $f(t) = \ln(t)$  conduz a estimadores dos vectores próprios com a menor variância assintótica, para um dado estimador de escala, enquanto que o desempenho dos estimadores dos valores próprios e das constantes de proporcionalidade é o mesmo do observado com a função identidade.

2. Para o modelo proporcional com dados normalmente distribuídos a variância assintótica do  $m$ -ésimo elemento do estimador de máxima verosimilhança da direcção comum  $\beta_j$  é  $\frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2}$  (Flury, 1986a, Flury, 1988, Capítulo 5). A variância assintótica dos estimadores dos valores próprios comuns com a aplicação da função  $f(t) = \ln(t)$  é por (5.49) e por (5.51) igual a

$$\text{ASVAR}(\hat{\beta}_{jm}) = \frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2} E_{\Phi} [\text{DIF}(Y, \sigma, \Phi)]^2. \quad (5.57)$$

Nestas condições, a eficiência relativa entre estes estimadores PP e os de máxima verosimilhança é

$$\text{Eff}(\hat{\beta}_{jm}) = \frac{\frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2}}{\frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2} E_{\Phi} [\text{DIF}(Y, \sigma, \Phi)]^2} = E_{\Phi} [\text{DIF}(Y, \sigma, \Phi)]^{-2}, \quad (5.58)$$

e portanto, não depende das constantes de proporcionalidade como acontece quando  $f(t) = t$ . Além disso, (5.58) no caso particular de  $\sigma^2(F) = \text{VAR}(F)$  atinge o valor 1. Este resultado mostra que estes novos estimadores PP de  $\beta_j$  têm assintoticamente o mesmo desempenho que os estimadores de máxima verosimilhança sob o modelo normal (se no índice de projecção se usar o estimador clássico da variância).

3. Com é evidente se  $(\lambda_{ij} = \lambda_{\ell j}), \forall_{i,\ell}$  então  $\rho_i = 1, \forall_i$ , pelo que estamos no caso de igualdade de matrizes e, à parte das localizações diferentes,  $V(\alpha)$  não depende de  $\alpha$ . Este problema é equivalente ao caso de  $k = 1$ , ou seja às *componentes principais* usuais e então conclui-se que, qualquer que seja  $f(\cdot)$  crescente e diferenciável, os estimadores obtidos são equivalentes (como não podia deixar de ser).
4. Admitindo, por exemplo, que  $\lambda_{1j} < \dots < \lambda_{kj}$  e que  $a_i < 1, \forall_i$  (para garantir que o denominador de (5.54) é positivo), é fácil concluir que  $V'(\alpha) > 0$  para  $\forall_{\alpha \geq 0}$  se  $a_1 < \dots < a_k$ , ou seja que também nestas condições  $f(t) = \ln(t)$  é óptima dentro da classe  $\mathcal{F}$ . De notar que por (5.54)  $V'(0)$  atinge o valor zero se e só se  $a_i = a_{\ell}, \forall_{i \neq \ell}$ , ou seja apenas no caso proporcional. Isto não implica, no entanto, que a função logaritmo seja óptima só para o modelo proporcional. De facto, como  $\alpha \geq 0$ , se tivermos  $V'(\alpha) > 0$  para  $\forall_{\alpha \geq 0}$  a função  $V(\alpha)$  atinge o mínimo em  $\alpha = 0$ .

## 5.6 Estudo de simulação

Desenvolveram-se dois estudos de simulação de Monte Carlo, um em dimensão 4 e outro em dimensão 2, para comparar o comportamento dos estimadores definidos em (5.3) com  $f(t) = \ln(t)$  e  $f(t) = t$ . O último caso corresponde à proposta de Boente e Orellana (2001) e foi também por nós estudado no Capítulo 4. As conclusões fundamentais do estudo presente encontram-se no artigo Boente, Pires e Rodrigues (2002b).

Em ambas as simulações foram considerados os estimadores PP resultantes da aplicação de  $f(t) = \ln(t)$  e  $f(t) = t$ . Usou-se a notação  $LPP_1$  e  $LPP_2$  para indicar os estimadores correspondentes à primeira função, enquanto que  $PP_1$  e  $PP_2$  referem os estimadores relacionados com a segunda opção. Em qualquer dos casos foram considerados dois estimadores robustos para a escala. O índice 1 indica que a escala é estimada pelo MAD (consistente sob o modelo normal) enquanto que o índice 2 é utilizado quando é empregue o estimador-M. A escolha do MAD deveu-se essencialmente à sua simplicidade de cálculo e, conseqüentemente, à diminuição do tempo de espera para a conclusão do processo de simulação. Quanto ao estimador-M de escala este, como referimos anteriormente (pág. 65), foi o considerado por Boente e Orellana (2001).

Com base nos resultados dum estudo prévio com um número reduzido de simulações, onde as direcções comuns foram estimadas recorrendo aos operadores máximo e mínimo, optou-se por utilizar neste estudo de Monte Carlo os estimadores consequência da maximização.

Para o cálculo computacional dos estimadores propostos foram utilizados os programas implementados em código MATLAB da autoria da Professora Graciela Boente e da Dra. Liliana Orellana. Desta forma evitámos que a comparação dos nossos resultados com os de Boente e Orellana (2001) fosse, eventualmente, influenciada pelo procedimento computacional. Além disso, e como não tínhamos à nossa disposição computadores adequados para cálculos intensivos, permitiu-nos usar o programa MATLAB que tem a vantagem de consumir menos tempo de CPU que a maior parte dos programas estatísticos. Este aspecto é mesmo crucial quando se trabalha com computadores pessoais com 256 MB de memória RAM. No entanto, também foram construídos novos códigos na linguagem do programa estatístico S-PLUS versão 2000. No Apêndice C apresentamos, como exemplo, a listagem do programa em S-PLUS que estima os parâmetros do modelo CPC através do estimador  $LPP_1$  construído com o operador máximo e que recorre a uma nova técnica para a pesquisa

das direcções. Na subsecção que se segue este novo método para a procura das direcções de projecção é apresentado detalhadamente.

### 5.6.1 Resultados da simulação em dimensão 4

Em dimensão  $p = 4$  foram considerados dois procedimentos distintos para o cálculo das várias estimativas “projection–pursuit”. O primeiro procedimento foi o considerado por Boente e Orellana (2001) e consiste numa adaptação da proposta de Croux e Ruiz–Gazen (1996) para a pesquisa das possíveis direcções em problemas com apenas uma população. Para o caso de  $k$  populações, o algoritmo pode ser resumido da forma seguinte,

- **PASSO 1.** Atribuir  $q = 1$ .
- **PASSO 2.** Para cada  $\mathbf{b} \in \mathbb{R}^p$ , tal que  $\|\mathbf{b}\| = 1$ , definir

$$r(\mathbf{b}) = \sum_{i=1}^k \tau_i f \{ s^2(\underline{\mathbf{X}}_i^T \mathbf{b}) \}.$$

Seja  $\mathbf{b}_0$  o valor que maximiza  $r(\mathbf{b})$ . Definir

$$\begin{aligned} \hat{\beta}_q &= \mathbf{b}_0, \\ \hat{\lambda}_{ij} &= s^2(\underline{\mathbf{X}}_i^T \mathbf{b}). \end{aligned}$$

- **PASSO 3.** Seja  $q = q + 1$ . Definir  $\underline{\mathbf{X}}_i^* = (\mathbf{I}_p - \mathbf{B}_{q-1} \mathbf{B}_{q-1}^T) \underline{\mathbf{X}}_i$ , para  $i = 1, \dots, k$  e onde  $\mathbf{B}_{q-1} = (\hat{\beta}_1, \dots, \hat{\beta}_{q-1})$ , para  $q = 2, \dots, p$ .
- **PASSO 4.** Se  $q \leq p$ , atribuir  $\underline{\mathbf{X}}_i = \underline{\mathbf{X}}_i^*$  para  $i = 1, \dots, k$  e voltar ao **PASSO 2**.
- **PASSO 5.** Ordenar os vectores próprios  $\hat{\beta}_1, \dots, \hat{\beta}_p$  de acordo com a ordem decrescente dos valores próprios da primeira população.

No **PASSO 2**, em vez de maximizar sobre todas as possíveis direcções, o que é na pratica impossível, usa-se o algoritmo de Croux e Ruiz–Gazen (1996) adaptado. Para isso, define-se  $\mathbf{P}_1 = \mathbf{I}_p$  e  $\mathbf{P}_q = (\mathbf{I}_p - \mathbf{B}_{q-1} \mathbf{B}_{q-1}^T)$ , ou seja a matriz de projecção no subespaço gerado pelos primeiros  $(q - 1)$  eixos. Considera-se  $\mathcal{A}_q = \{\mathbf{P}_q(\mathbf{X}_{ij} - \hat{\mu}_i), j = 1, \dots, n_i, i = 1, \dots, k\}$ , onde  $\hat{\mu}_i$  representa uma estimativa de

localização com *alto ponto de ruptura* para a  $i$ -ésima população (por exemplo o SDE de localização). Sendo  $M_q$  o número de vectores não nulos em  $\mathcal{A}_q$  e  $\mathbf{a}_\ell$  os vectores de  $\mathcal{A}_q$  com  $\ell = 1 \dots, M_q$ , a pesquisa será apenas efectuada nas direcções  $\frac{\mathbf{a}_\ell}{\|\mathbf{a}_\ell\|}$ , para  $\ell = 1 \dots, M_q$ . Para este processo de procura utilizou-se a sigla DD, construída a partir das iniciais de “data directions”, porque as possíveis direcções comuns são procuradas nas direcções dos dados. Este método de pesquisa está condicionado ao número de observações, o que pode ser problemático em casos com reduzido número de observações, especialmente em dimensões elevadas. Na tentativa de contornar este problema, foi considerada uma nova alternativa para a pesquisa das direcções comuns. Com o novo procedimento, denotado por RD, iniciais da designação “random directions”, as direcções comuns são procuradas fazendo as alterações que se seguem ao processo DD,

- i) geram-se aleatoriamente 1000 direcções através de  $\mathbf{Z}_i \sim N(\mathbf{0}_p, \mathbf{I}_p)$ ,
- ii) considera-se  $\mathcal{A}_q = \{\mathbf{P}_q(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}), j = 1 \dots, 1000\}$ , onde  $\hat{\boldsymbol{\mu}}$  é o SDE de localização de  $(\mathbf{Z}_1, \dots, \mathbf{Z}_{1000})$ .

Note-se que, na alternativa RD são criadas 1000 direcções uniformemente distribuídas sobre a esfera unitária de  $\mathbb{R}^p$ . Isto é consequência do resultado de Kariya e Eaton (1977) onde os autores provam que um vector aleatório  $\mathbf{X}$ , com distribuição esférica, pode ser factorizado como  $\mathbf{X} = \mathbf{R}\mathbf{U}$  em que  $\mathbf{R} = \|\mathbf{X}\|$  é estocasticamente independente de  $\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$  e  $\mathbf{U}$  é uniformemente distribuído sobre a esfera unitária de  $\mathbb{R}^p$ .

No estudo desenvolvido foram considerados os três modelos que se seguem:

- **Modelo 1.**  $k = 2$  populações com  $\boldsymbol{\Sigma}_1 = \text{diag}(4, 3, 2, 1)$  e  $\boldsymbol{\Sigma}_2 = \text{diag}(6, 8, 2.5, 5)$ . Os funcionais com  $f(t) = \text{id}(t)$  são consistentes porque  $\boldsymbol{\Sigma}_{\text{id}} = \sum_{i=1}^k \tau_i \boldsymbol{\Sigma}_i = \text{diag}(5, 5.5, 2.25, 3)$ . Esperam-se problemas em relação a  $f(t) = \ln(t)$  porque  $\mathbf{N}_{\ln}$  é tal que  $\mathbf{N}_{\ln}(1, 2) < 0$  e  $\mathbf{N}_{\ln}(3, 4) < 0$  (as restantes entradas são não negativas). Esta foi a situação estudada por Boente e Orellana (2001).
- **Modelo 2.**  $k = 2$  populações com matrizes de dispersão proporcionais tais que  $\boldsymbol{\Sigma}_1 = \text{diag}(4, 3, 2, 1)$  e  $\boldsymbol{\Sigma}_2 = 4 \boldsymbol{\Sigma}_1$ . Os funcionais com  $f(t) = \text{id}(t)$  e com  $f(t) = \ln(t)$  são consistentes porque se mantém a ordem dos valores próprios nas duas populações.
- **Modelo 3.**  $k = 3$  populações com matrizes de dispersão dadas por  $\boldsymbol{\Sigma}_1 = \text{diag}(8, 4, 2, 1)$ ,  $\boldsymbol{\Sigma}_2 = 5 \boldsymbol{\Sigma}_1$  e  $\boldsymbol{\Sigma}_3 = \text{diag}(16, 12, 8, 4)$ . Os funcionais com  $f(t) =$



$\text{id}(t)$  e com  $f(t) = \ln(t)$  são consistentes porque se mantêm a ordem dos valores próprios nas três populações.

Em todos os modelos, foram efectuadas 1000 replicações gerando-se  $k$  amostras independentes com dimensão  $n_i = n = 100$ ,  $i = 1, 2$ . Os verdadeiros eixos principais são os eixos do sistema cartesiano, dados pelos vectores unitários  $\mathbf{e}_j$ . Os vectores próprios foram ordenados de acordo com a ordem decrescente dos valores próprios da primeira população e portanto  $\beta_j = \mathbf{e}_j$ .

No estudo desenvolvido foram gerados dados normalmente distribuídos, que se referenciam por  $C_0$ , enquanto que  $C_{1,\varepsilon}$  e  $C_{2,\varepsilon}$  representam os dois tipos de dados contaminados apresentados a seguir:

- $C_{1,\varepsilon}$ :  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in}$  são i.i.d.  $0.9N(\mathbf{0}_4, \Sigma_i) + 0.1N(\mathbf{0}_4, 9\Sigma_i)$ , fixando-se  $\varepsilon = 10\%$ .
- $C_{2,\varepsilon}$ :  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in}$  são i.i.d.  $(1-\varepsilon)N(\mathbf{0}_4, \Sigma_i) + \varepsilon N(\boldsymbol{\mu}, \Sigma_i)$  com  $\boldsymbol{\mu} = (0, 0, 0, 10)^T$ , para  $\varepsilon = 5\%$  e  $\varepsilon = 10\%$ .

Estes cenários são os mesmos que foram considerados por Boente e Orellana (2001) para avaliar o comportamento dos estimadores  $\text{PP}_1$  e  $\text{PP}_2$  no Modelo 1. Por uma questão de simplicidade e seguindo os trabalhos de Orellana (1999) e Boente e Orellana (2001), apresentamos apenas os resultados referentes à estimação dos vectores próprios comuns e dos valores próprios para o primeiro grupo.

As Tabelas 5.1 e 5.2 dão-nos medidas sumárias relativas à estimação dos vectores próprios comuns para o Modelo 1, indicando respectivamente, a mediana e a média das distâncias entre o vector estimado e o verdadeiro vector próprio, i.e.,  $\|\hat{\beta}_j - \beta_j\|^2$ . Esta distância permite também determinar o ângulo,  $\hat{\theta}_j$ , entre a  $j$ -ésima direcção estimada e a verdadeira direcção fazendo  $\cos(\hat{\theta}_j) = 1 - \|\hat{\beta}_j - \beta_j\|^2 / 2$ . É fácil verificar que o último resultado é consequência da igualdade  $\|\hat{\beta}_j - \beta_j\|^2 = \|\hat{\beta}_j\|^2 + \|\beta_j\|^2 - 2\|\hat{\beta}_j\|\|\beta_j\|\cos\hat{\theta}_j$ .

As Tabelas 5.3 e 5.4 sumarizam os resultados para o Modelo 2, enquanto que as Tabelas 5.5 e 5.6 referem-se ao último modelo, Modelo 3.

Para o Modelo 3, por uma questão de economia de tempo, só observamos o comportamento dos estimadores sob o efeito da contaminação  $C_{2,\varepsilon}$ . Relembre-se que a contaminação  $C_{1,\varepsilon}$  só influencia a variância dos eixos principais de cada direcção não alterando as direcções dos eixos e o objectivo deste estudo é avaliar o efeito de  $f(t)$  na estimação dos eixos principais. A nossa opção é justificada pelo facto da escolha da função  $f(t)$  não influenciar, assintoticamente, a estimação dos valores próprios (Teorema 5.2).

Representando por  $\text{ASE}_j(f)$  a média assintótica de  $n\|\widehat{\beta}_j - \beta_j\|^2$  é fácil verificar que  $\text{ASE}_j(f) = \sum_{m \neq j} \text{ASVAR}(\widehat{\beta}_{jm})$ , uma vez que  $\text{ASVAR}(\widehat{\beta}_{jj}) = 0$ . Assim, sob as condições estabelecidas na Proposição 5.1, esta expressão atingirá o seu mínimo quando  $f(t) = \ln(t)$ . Referenciando por  $V^*(f)$  a função de  $f$  que pré-multiplica o operador valor esperado em (5.49), a aplicação da função  $f(t) = t$  em (5.49) conduz a

$$V^*(\text{id}) = \frac{\lambda_j \lambda_m}{(\lambda_m - \lambda_j)^2} \frac{\sum_{i=1}^k \tau_i \rho_i^2}{\left\{ \sum_{i=1}^k \tau_i \rho_i \right\}^2}, \quad \text{para } m \neq j. \quad (5.59)$$

Por outro lado, o emprego de  $f(t) = \ln(t)$  origina

$$V^*(\ln) = \frac{\lambda_j \lambda_m}{(\lambda_m - \lambda_j)^2}, \quad \text{para } m \neq j. \quad (5.60)$$

Por conseguinte, o rácio  $R_p = \frac{\text{ASE}_j(\text{id})}{\text{ASE}_j(\ln)}$  iguala  $\left( \sum_{i=1}^k \tau_i \rho_i \right)^{-2} \sum_{i=1}^k \tau_i \rho_i^2$ . Assim, sob  $C_0$ , e para o modelo proporcional aqui considerado, tem-se  $R_4 = \frac{34}{25} = 1.36$ . Note-se que os valores da Tabela 5.4, para o estimador-M de escala, traduzem uma boa aproximação do valor de  $R_4$ . Por exemplo, para o caso da primeira direcção principal observa-se para valor estimado de  $R_4$  aproximadamente 1.27. Os resultados desta tabela mostram claramente a vantagem de  $f(t) = \ln(t)$  face à função identidade na estimação das direcções comuns. Refira-se ainda que para os outros modelos também é possível calcular valores do mesmo tipo, embora dependentes do vector próprio em causa.

As Tabelas 5.7, 5.8 e 5.9 reportam as médias, desvios padrões e erros quadráticos médios das estimativas dos valores próprios de  $\Sigma_1$  para o Modelo 1, Modelo 2 e Modelo 3, respectivamente.

Os “melhores” resultados para a estimação dos valores próprios, sob contaminação, são obtidos com as estimativas baseadas no estimador-M de escala, ver Tabelas 5.7, 5.8 e 5.9. Contrariamente às nossas expectativas, não se observou uma grande melhoria quando se pesquisa em 1000 direcções aleatórias em vez de tomar as direcções das observações normalizadas. A pesquisa em direcções aleatórias poderá, eventualmente, ser uma boa alternativa para amostras de mais baixas dimensões. A fraca eficiência das estimativas dos valores próprios, em ambos os procedimentos “projection-pursuit”, poderá estar relacionada com a baixa eficiência do MAD e do estimador-M de escala escolhido. Melhores resultados seriam provavelmente obtidos aplicando estimadores de escala mais eficientes. Uma alternativa seria utilizar um estimador-M ponderado com uma outra função  $\chi$  ou, seguindo a sugestão de

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.4157	0.3654	0.4166	0.3618	0.3909	0.3580	0.4063	0.3583
4	$C_{1,0.10}$	0.4125	0.3672	0.4002	0.3527	0.4264	0.3607	0.4078	0.3656
	$C_{2,0.05}$	0.4678	0.4582	0.4863	0.4712	0.4585	0.4409	0.4783	0.4595
	$C_{2,0.10}$	0.5411	0.5365	0.5832	0.5397	0.5366	0.4986	0.5692	0.5710
	$C_0$	0.4975	0.4029	0.4781	0.3931	0.4491	0.3913	0.4359	0.3931
3	$C_{1,0.10}$	0.5080	0.4041	0.4815	0.3931	0.4808	0.3919	0.4708	0.4155
	$C_{2,0.05}$	0.6230	0.5492	0.6440	0.5299	0.5475	0.5095	0.5505	0.4909
	$C_{2,0.10}$	0.7670	0.7243	0.7367	0.7038	0.6778	0.6630	0.6618	0.7002
	$C_0$	0.3956	0.2363	0.4378	0.2031	0.4857	0.4028	0.5057	0.3968
2	$C_{1,0.10}$	0.3764	0.2255	0.4139	0.2322	0.4422	0.4045	0.4830	0.3906
	$C_{2,0.05}$	0.8699	0.8205	0.9616	0.7572	0.7349	0.6776	0.7807	0.6558
	$C_{2,0.10}$	1.4109	1.5810	1.4301	1.6368	1.2473	1.3623	1.3024	1.4167
	$C_0$	0.3625	0.2814	0.4058	0.2685	0.4087	0.3606	0.4114	0.3780
1	$C_{1,0.10}$	0.3519	0.2679	0.4106	0.2790	0.3807	0.3815	0.4414	0.3489
	$C_{2,0.05}$	0.6790	0.6972	0.7988	0.7259	0.6847	0.6772	0.7050	0.6744
	$C_{2,0.10}$	1.3291	1.5579	1.2906	1.5869	1.1359	1.3248	1.1940	1.4114

**Tabela 5.1:** Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 1.

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.5813	0.5327	0.5743	0.5479	0.5552	0.5279	0.5611	0.5437
4	$C_{1,0.10}$	0.5979	0.5428	0.5755	0.5309	0.5918	0.5430	0.5695	0.5372
	$C_{2,0.05}$	0.6132	0.6191	0.6395	0.6142	0.6195	0.6089	0.6266	0.5864
	$C_{2,0.10}$	0.6836	0.6901	0.7231	0.6989	0.6895	0.6648	0.6988	0.6987
	$C_0$	0.7120	0.6203	0.6986	0.6184	0.6453	0.5774	0.6195	0.5919
3	$C_{1,0.10}$	0.7135	0.6221	0.7061	0.6111	0.6654	0.5857	0.6329	0.5836
	$C_{2,0.05}$	0.7771	0.7373	0.7848	0.7174	0.7234	0.6780	0.7121	0.6426
	$C_{2,0.10}$	0.8878	0.8551	0.8829	0.8617	0.8106	0.7822	0.8105	0.8245
	$C_0$	0.6777	0.5219	0.7062	0.4685	0.6429	0.5108	0.6350	0.4940
2	$C_{1,0.10}$	0.6572	0.4750	0.6796	0.4910	0.6037	0.5086	0.6294	0.5001
	$C_{2,0.05}$	0.9127	0.8936	0.9453	0.8663	0.8668	0.8031	0.8652	0.7828
	$C_{2,0.10}$	1.1987	1.3167	1.1975	1.3800	1.1474	1.2045	1.1756	1.2506
	$C_0$	0.5677	0.4703	0.6100	0.4498	0.5445	0.4636	0.5643	0.4610
1	$C_{1,0.10}$	0.5426	0.4439	0.5803	0.4454	0.5216	0.4654	0.5616	0.4582
	$C_{2,0.05}$	0.8600	0.8909	0.9169	0.8898	0.8165	0.8020	0.8279	0.8112
	$C_{2,0.10}$	1.2129	1.3491	1.2061	1.3968	1.1371	1.2428	1.1813	1.2935

**Tabela 5.2:** Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 1.

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.2833	0.2046	0.3409	0.2002	0.2264	0.1516	0.2298	0.1739
4	$C_{1,0.10}$	0.2832	0.1986	0.2981	0.1980	0.2089	0.1544	0.2232	0.1886
	$C_{2,0.05}$	0.2932	0.2471	0.3570	0.2360	0.2473	0.1983	0.2809	0.1910
	$C_{2,0.10}$	0.3727	0.3421	0.4239	0.3768	0.3143	0.2800	0.3600	0.2762
	$C_0$	0.5535	0.4069	0.5747	0.3951	0.3874	0.2558	0.3973	0.2860
3	$C_{1,0.10}$	0.5140	0.3779	0.5180	0.3931	0.3704	0.2681	0.3930	0.2842
	$C_{2,0.05}$	0.5893	0.4492	0.6428	0.4574	0.4204	0.3388	0.4711	0.3348
	$C_{2,0.10}$	0.7244	0.6642	0.7730	0.6559	0.5634	0.5024	0.5655	0.4794
	$C_0$	0.3443	0.2060	0.3405	0.1940	0.2454	0.1532	0.2437	0.1595
2	$C_{1,0.10}$	0.3286	0.2128	0.3486	0.1959	0.2522	0.1643	0.2561	0.1586
	$C_{2,0.05}$	0.5070	0.3766	0.5086	0.3734	0.3446	0.2759	0.3612	0.2903
	$C_{2,0.10}$	0.7076	0.7313	0.8000	0.7574	0.5834	0.5520	0.6173	0.6001
	$C_0$	0.1268	0.0650	0.1293	0.0646	0.1056	0.0554	0.1078	0.0522
1	$C_{1,0.10}$	0.1246	0.0712	0.1319	0.0693	0.1046	0.0562	0.1117	0.0581
	$C_{2,0.05}$	0.2740	0.2458	0.2809	0.2541	0.2460	0.2427	0.2662	0.2500
	$C_{2,0.10}$	0.5271	0.5820	0.6036	0.6211	0.4875	0.5601	0.5427	0.5960

**Tabela 5.3:** Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 2.

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.4977	0.4147	0.5240	0.4137	0.4162	0.3263	0.4130	0.3244
4	$C_{1,0.10}$	0.4881	0.4075	0.4730	0.4140	0.4081	0.3276	0.4251	0.3214
	$C_{2,0.05}$	0.5107	0.4584	0.5426	0.4444	0.4168	0.3767	0.4446	0.3526
	$C_{2,0.10}$	0.5910	0.5501	0.6225	0.5575	0.4882	0.4502	0.5153	0.4505
	$C_0$	0.7212	0.5926	0.7218	0.5848	0.5800	0.4588	0.5589	0.4473
3	$C_{1,0.10}$	0.6845	0.5798	0.6962	0.5838	0.5499	0.4526	0.5559	0.4486
	$C_{2,0.05}$	0.7525	0.6532	0.7777	0.6444	0.6057	0.5180	0.6243	0.5045
	$C_{2,0.10}$	0.8417	0.8107	0.8814	0.8030	0.7250	0.6775	0.7243	0.6617
	$C_0$	0.5466	0.3849	0.5312	0.3706	0.3942	0.2707	0.3666	0.2677
2	$C_{1,0.10}$	0.5278	0.3861	0.5464	0.3677	0.3777	0.2712	0.3643	0.2683
	$C_{2,0.05}$	0.6885	0.5617	0.6762	0.5516	0.5173	0.3996	0.5225	0.4337
	$C_{2,0.10}$	0.8449	0.8377	0.8829	0.8514	0.7256	0.7334	0.7530	0.7355
	$C_0$	0.2211	0.1182	0.2106	0.1150	0.1611	0.0863	0.1532	0.0822
1	$C_{1,0.10}$	0.2300	0.1244	0.2322	0.1190	0.1662	0.0915	0.1631	0.0900
	$C_{2,0.05}$	0.4279	0.3535	0.4173	0.3682	0.3461	0.2943	0.3526	0.3211
	$C_{2,0.10}$	0.6488	0.7111	0.7092	0.7404	0.6030	0.6684	0.6535	0.6997

**Tabela 5.4:** Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 2.

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.0867	0.0474	0.1212	0.0472	0.0808	0.0434	0.0984	0.0455
8	$C_{2,0.05}$	0.0930	0.0531	0.1236	0.0622	0.0908	0.0554	0.1233	0.0865
	$C_{2,0.10}$	0.1090	0.0690	0.1321	0.0889	0.1084	0.0806	0.1321	0.1062
	$C_0$	0.1631	0.0804	0.1874	0.0838	0.1289	0.0710	0.1536	0.0775
4	$C_{2,0.05}$	0.1867	0.1165	0.2236	0.1269	0.1716	0.1155	0.1974	0.1226
	$C_{2,0.10}$	0.2514	0.1972	0.3027	0.2111	0.2528	0.2045	0.2750	0.2316
	$C_0$	0.1674	0.0800	0.1832	0.0878	0.1379	0.0678	0.1430	0.0698
2	$C_{2,0.05}$	0.2667	0.2029	0.2989	0.2014	0.2311	0.1828	0.2677	0.1968
	$C_{2,0.10}$	0.4936	0.4494	0.5561	0.5226	0.4347	0.4729	0.5252	0.5400
	$C_0$	0.0878	0.0396	0.0947	0.0416	0.0720	0.0311	0.0814	0.0347
1	$C_{2,0.05}$	0.2001	0.1686	0.2195	0.1720	0.1916	0.1805	0.2177	0.1896
	$C_{2,0.10}$	0.4197	0.4458	0.4657	0.5080	0.4282	0.5018	0.5237	0.5801

**Tabela 5.5:** Mediana do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 3.

$\lambda$		PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$	0.1659	0.0979	0.1823	0.1204	0.1353	0.0842	0.1555	0.0879
8	$C_{2,0.05}$	0.1744	0.1076	0.2008	0.1157	0.1535	0.0976	0.1640	0.1103
	$C_{2,0.10}$	0.2083	0.1393	0.2298	0.1388	0.1731	0.1286	0.2011	0.1495
	$C_0$	0.2916	0.1559	0.3133	0.1760	0.2116	0.1248	0.2296	0.1209
4	$C_{2,0.05}$	0.3241	0.1990	0.3666	0.2038	0.2620	0.1675	0.2714	0.1748
	$C_{2,0.10}$	0.4121	0.3169	0.4634	0.3215	0.3425	0.2770	0.3741	0.3170
	$C_0$	0.3120	0.1596	0.3369	0.1629	0.2348	0.1079	0.2316	0.1181
2	$C_{2,0.05}$	0.4463	0.3326	0.4855	0.3360	0.3596	0.2742	0.3845	0.2905
	$C_{2,0.10}$	0.6469	0.5852	0.7109	0.6647	0.6075	0.6174	0.6804	0.7051
	$C_0$	0.1871	0.0942	0.2011	0.0933	0.1489	0.0592	0.1483	0.0699
1	$C_{2,0.05}$	0.3425	0.2858	0.3725	0.2899	0.2984	0.2649	0.3270	0.2767
	$C_{2,0.10}$	0.5629	0.5512	0.6088	0.6356	0.5614	0.6249	0.6488	0.7044

**Tabela 5.6:** Média do quadrado da distância entre a direcção principal comum estimada e o verdadeiro eixo principal relacionado com o valor próprio  $\lambda$  da primeira população, Modelo 3.

Rousseeuw e Croux (1994), considerar só 1 ou 2 passos na ponderação dos observações.

Para analisar melhor os resultados referentes aos vectores próprios, estimou-se a densidade da variável aleatória  $\cos(\hat{\theta}_1)$  através do estimador do núcleo (“kernel esti-

Estimativas de  $\lambda_{14} = 4$ 

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	4.2507	4.0159	4.3187	4.0210	4.5268	4.1864	4.6435	4.2152	5.0189	4.7612	5.0660	4.7723	5.3760	4.9851	5.4338	4.9987
DP	0.9283	0.7401	0.9079	0.7369	0.8114	0.6788	0.8203	0.6811	1.0748	0.8977	1.0578	0.9113	1.0111	0.8676	0.9867	0.8404
EQM	0.9247	0.5480	0.9259	0.5435	0.9363	0.4956	1.0874	0.5130	2.1945	1.3859	2.2565	1.4274	2.9178	1.7242	3.0314	1.7048
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	4.4626	4.1910	4.5065	4.2235	4.7527	4.3781	4.8727	4.4510	4.8875	4.6880	4.9462	4.6846	5.2014	4.8883	5.3355	4.9018
DP	0.9666	0.7598	0.9390	0.7719	0.8574	0.7225	0.8592	0.7029	1.0493	0.9607	1.0662	0.8914	0.9435	0.8811	0.9752	0.8439
EQM	1.1485	0.6138	1.1233	0.6458	1.3022	0.6650	1.5006	0.6977	1.8895	1.3968	2.0330	1.2638	2.3351	1.5661	2.7363	1.5263

Estimativas de  $\lambda_{13} = 3$ 

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	3.0987	3.0123	3.1372	3.0328	3.4153	3.1846	3.4678	3.1927	3.6561	3.5881	3.6901	3.6184	4.0223	3.7816	4.0696	3.8113
DP	0.6463	0.5335	0.6264	0.5226	0.6083	0.5337	0.6137	0.5315	0.7485	0.6586	0.7602	0.6434	0.7368	0.6416	0.7174	0.6492
EQM	0.4274	0.2848	0.4112	0.2742	0.5427	0.3189	0.5957	0.3197	0.9912	0.7800	1.0546	0.7967	1.5890	1.0231	1.6600	1.0802
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	3.3146	3.2235	3.3738	3.2572	3.6369	3.4086	3.7442	3.4507	3.7223	3.6871	3.7787	3.7310	4.0646	3.8947	4.1734	3.9473
DP	0.6464	0.5565	0.6422	0.5678	0.6499	0.5628	0.6262	0.5599	0.7363	0.6870	0.7252	0.6973	0.7136	0.6924	0.7168	0.6790
EQM	0.5169	0.3597	0.5523	0.3886	0.8284	0.4838	0.9464	0.5168	1.0644	0.9445	1.1329	1.0212	1.6438	1.2806	1.8919	1.3593

Estimativas de  $\lambda_{12} = 2$ 

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	2.0975	1.9703	2.0992	1.9914	2.1376	1.9221	2.1818	1.9354	2.4406	2.3483	2.4605	2.3595	2.5216	2.2859	2.5524	2.3052
DP	0.4373	0.3732	0.4362	0.3749	0.4534	0.3682	0.4541	0.3627	0.5330	0.4483	0.5344	0.4589	0.5367	0.4515	0.5534	0.4608
EQM	0.2007	0.1402	0.2001	0.1406	0.2245	0.1416	0.2393	0.1358	0.4784	0.3225	0.4978	0.3399	0.5604	0.2857	0.6117	0.3056
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	2.3912	2.3214	2.4186	2.3216	2.4734	2.2891	2.5158	2.2955	2.8370	2.8960	2.9013	2.9217	2.9845	2.9043	3.0350	2.9352
DP	0.5011	0.4425	0.4942	0.4419	0.5352	0.4415	0.5353	0.4405	0.6230	0.6127	0.6141	0.6302	0.6517	0.6307	0.6550	0.6481
EQM	0.4043	0.2992	0.4196	0.2988	0.5107	0.2785	0.5529	0.2814	1.0894	1.1790	1.1903	1.2475	1.3949	1.2163	1.5014	1.2954

Estimativas de  $\lambda_{11} = 1$ 

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	1.4424	1.3821	1.4788	1.3942	1.4102	1.3339	1.4246	1.3513	1.6695	1.6339	1.7308	1.6563	1.6473	1.5979	1.7007	1.6096
DP	0.3658	0.3164	0.3672	0.3148	0.3546	0.2806	0.3394	0.2845	0.4495	0.3917	0.4390	0.3895	0.4160	0.3549	0.4305	0.3496
EQM	0.3297	0.2462	0.3643	0.2546	0.2941	0.1903	0.2957	0.2045	0.6508	0.5556	0.7273	0.5830	0.5924	0.4838	0.6767	0.4942
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	1.7654	1.7884	1.8246	1.8185	1.7389	1.7533	1.7838	1.7829	2.0903	2.1274	2.1363	2.1490	2.1137	2.1493	2.1571	2.1586
DP	0.4241	0.3737	0.4359	0.3901	0.4165	0.3687	0.4262	0.3643	0.4897	0.4334	0.4871	0.4492	0.4956	0.4458	0.5209	0.4650
EQM	0.7662	0.7619	0.8707	0.8228	0.7201	0.7040	0.7966	0.7463	1.4298	1.4600	1.5298	1.5233	1.4871	1.5210	1.6115	1.5599

**Tabela 5.7:** Estimação dos valores próprios de  $\Sigma_1$ , Modelo 1.

mator”). Relembre-se que o estimador do núcleo, com núcleo  $K$ , é definido como

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

onde  $h$  é a largura da banda (“bandwidth”). A função núcleo  $K$  deve ser tal que  $\int_{-\infty}^{+\infty} K(x) dx = 1$ . Mais pormenores encontram-se, por exemplo, na referência Silverman (1986, Capítulo 3).

Neste estudo considerou-se, seguindo Boente e Orellana (2001), para núcleo a densidade da distribuição normal e tomou-se para largura da banda o valor 0.3. As Figuras 5.1 a 5.3 apresentam as estimativas obtidas a partir de 1000 réplicas. As linhas a preto correspondem à situação  $C_0$ , enquanto que as linhas a azul e a ver-

Estimativas de  $\lambda_{14} = 4$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	4.1561	4.0567	4.1736	4.0475	4.6734	4.3093	4.7302	4.3335	4.8819	4.7954	4.8978	4.7737	5.4895	5.1330	5.5513	5.1420
DP	0.8642	0.726	0.8847	0.7259	0.8212	0.6999	0.8148	0.6983	1.0484	0.9135	1.0525	0.8733	1.0362	0.8826	1.0121	0.8508
EQM	0.7712	0.5315	0.8129	0.5292	1.1283	0.5856	1.1975	0.5989	1.8777	1.4678	1.9147	1.3618	3.2945	2.0638	3.4333	2.0293
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	4.3191	4.1686	4.3356	4.2136	4.8844	4.5225	4.9843	4.5843	4.6713	4.5713	4.6797	4.5953	5.2958	4.9920	5.4353	5.0455
DP	0.8861	0.7478	0.8850	0.7492	0.8614	0.7258	0.8684	0.7095	0.9979	0.8907	1.0012	0.8515	0.9540	0.8783	0.9625	0.8352
EQM	0.8871	0.5877	0.8960	0.6070	1.5250	0.8001	1.7240	0.8451	1.4468	1.1201	1.4648	1.0798	2.5910	1.7564	2.9885	1.7916

Estimativas de  $\lambda_{13} = 3$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	3.0264	2.9773	3.0358	2.9918	3.3956	3.1349	3.4586	3.1496	3.5316	3.5306	3.5955	3.5532	4.0036	3.7229	4.0667	3.7441
DP	0.5791	0.5129	0.5946	0.5156	0.6044	0.5264	0.6072	0.5282	0.7070	0.6208	0.7281	0.6576	0.7349	0.6420	0.7297	0.6440
EQM	0.3361	0.2636	0.3549	0.2659	0.5220	0.2953	0.5792	0.3014	0.7827	0.6672	0.8851	0.7388	1.5482	0.9353	1.6715	0.9690
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	3.2034	3.1415	3.2115	3.1492	3.6065	3.3743	3.7023	3.3997	3.5249	3.5245	3.5795	3.5669	4.0018	3.8217	4.1279	3.8710
DP	0.6155	0.5411	0.6066	0.5320	0.6342	0.5650	0.6290	0.5482	0.6900	0.6358	0.6673	0.6487	0.7018	0.6469	0.7132	0.6724
EQM	0.4203	0.3129	0.4127	0.3053	0.7704	0.4594	0.8893	0.4604	0.7519	0.6796	0.7815	0.7424	1.4971	1.0944	1.7822	1.2116

Estimativas de  $\lambda_{12} = 2$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	2.1478	2.1082	2.1807	2.1148	2.2873	2.1025	2.3233	2.1242	2.5082	2.5053	2.5353	2.5112	2.6924	2.5011	2.7047	2.5075
DP	0.4506	0.4097	0.4555	0.4093	0.4400	0.3835	0.4593	0.3833	0.5520	0.5097	0.5355	0.4962	0.5553	0.4899	0.5446	0.4732
EQM	0.2249	0.1796	0.2401	0.1807	0.2763	0.1576	0.3156	0.1624	0.5633	0.5153	0.5736	0.5078	0.7882	0.4914	0.7938	0.4818
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	2.3280	2.3056	2.3638	2.3152	2.5105	2.3586	2.5656	2.3637	2.6818	2.7490	2.7278	2.7773	2.9132	2.8468	2.9862	2.8615
DP	0.4760	0.4329	0.4813	0.4374	0.4953	0.4403	0.5030	0.4285	0.5684	0.5623	0.5525	0.5567	0.6012	0.6107	0.5899	0.5981
EQM	0.3343	0.2809	0.3641	0.2908	0.5062	0.3226	0.5733	0.3161	0.7884	0.8777	0.8355	0.9147	1.1963	1.0907	1.3215	1.1005

Estimativas de  $\lambda_{11} = 1$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$								$C_{1,0.10}$							
Média	1.2088	1.1144	1.2253	1.1257	1.1410	1.0615	1.1591	1.0603	1.4290	1.3104	1.4323	1.3335	1.3602	1.2563	1.3669	1.2653
DP	0.3130	0.2420	0.3225	0.2453	0.3017	0.2185	0.3009	0.2193	0.3850	0.2973	0.3759	0.2908	0.3677	0.2691	0.3542	0.2734
EQM	0.1416	0.0717	0.1549	0.0760	0.1109	0.0515	0.1159	0.0518	0.3325	0.1848	0.3284	0.1959	0.2650	0.1382	0.2603	0.1452
	$C_{2,0.05}$								$C_{2,0.10}$							
Mean	1.5019	1.5120	1.5528	1.5351	1.4629	1.4455	1.5024	1.4819	1.9563	2.0887	2.0125	2.1198	1.9121	2.0235	1.9893	2.0777
SD	0.3965	0.3689	0.4079	0.3714	0.4036	0.3515	0.4130	0.3733	0.5268	0.5068	0.5188	0.5139	0.5478	0.5276	0.5475	0.5372
MSE	0.4093	0.3985	0.4723	0.4246	0.3774	0.3222	0.4233	0.3718	1.1930	1.4432	1.2953	1.5194	1.1328	1.3269	1.2794	1.4512

**Tabela 5.8:** Estimação dos valores próprios de  $\Sigma_1$ , Modelo 2.

melhor referem-se às contaminações assimétricas geradas de acordo com  $C_{2,0.05}$  e  $C_{2,0.10}$ , respectivamente.

Sob a contaminação  $C_{2,0.10}$ , as estimativas de “projection–pursuit” apresentam um desempenho para o Modelo 1 diferente do observado nos Modelos 2 e 3. Nota-se uma enorme sensibilidade em todas as propostas para o Modelo 1, enquanto que nos Modelos 2 e 3 as estimativas apresentam um comportamento muito mais estável mesmo quando sujeitas à contaminação. Note-se que para o Modelo 1 a matriz de dispersão combinada é  $\Sigma_{id} = \text{diag}(5, 5.5, 2.25, 3)$  e, como consequência, as estimativas “projection–pursuit” com o procedimento  $f(t) = t$  vão “quebrar” com 10% de contaminação. Isto acontece mesmo quando estimamos a escala com o MAD, como foi salientado por Boente e Orellana (2001). Além disso, como os maiores valores próprios de  $\Sigma_{id}$  são próximos a dimensão amostral não é suficiente para os

Estimativas de  $\lambda_{14} = 8$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$															
Média	7.7778	7.8214	7.7389	7.7742	8.4729	8.0732	8.4621	8.0216								
DP	1.8446	1.5978	1.8248	1.5948	1.7975	1.5609	1.7789	1.5497								
EQM	3.4518	2.5848	3.3983	2.5943	3.4547	2.4419	3.3781	2.4021								
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	7.8649	7.8709	7.8481	7.8580	8.6465	8.1985	8.6792	8.1974	8.0854	8.0392	8.1072	8.0441	8.9656	8.5314	9.0546	8.5807
DP	1.8281	1.6141	1.8091	1.5847	1.8164	1.5637	1.8006	1.5479	1.9083	1.6587	1.8861	1.6380	1.8916	1.6098	1.8662	1.5807
EQM	3.3602	2.6222	3.2960	2.5314	3.7179	2.4846	3.7041	2.4351	3.6490	2.7528	3.5689	2.6849	4.5115	2.8742	4.5961	2.8362

Estimativas de  $\lambda_{13} = 4$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$															
Média	4.1898	4.0789	4.2130	4.1103	4.5043	4.1683	4.5742	4.1809								
DP	0.9783	0.8462	0.9829	0.8616	0.9503	0.7968	0.9442	0.8128								
EQM	0.9930	0.7223	1.0116	0.7544	1.1578	0.6633	1.2214	0.6935								
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	4.3119	4.2111	4.3628	4.2081	4.7294	4.3992	4.8229	4.4200	4.6166	4.5646	4.6858	4.5804	5.1400	4.8751	5.2499	4.9288
DP	0.9951	0.8533	1.0040	0.8733	0.9762	0.8273	1.0172	0.8502	1.0662	0.9466	1.0592	0.9528	1.0510	0.9750	1.1259	0.9810
EQM	1.0876	0.7728	1.1398	0.8060	1.4855	0.8439	1.7125	0.8993	3.6490	1.2152	1.5928	1.2450	2.4055	1.7173	2.8313	1.8260

Estimativas de  $\lambda_{12} = 2$

	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$															
Média	2.2520	2.1179	2.2851	2.1454	2.3613	2.1182	2.3880	2.1503								
DP	0.5546	0.4549	0.5511	0.4536	0.5367	0.4147	0.5207	0.4167								
EQM	0.3712	0.2209	0.3851	0.2269	0.4187	0.1860	0.4219	0.1962								
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	2.4536	2.3138	2.5249	2.3529	2.6221	2.3915	2.6709	2.4154	2.8445	2.8216	2.9966	2.8817	3.1222	2.9691	3.2065	3.0505
DP	0.5851	0.4906	0.5981	0.4850	0.6036	0.4827	0.5784	0.4760	0.7320	0.6985	0.7479	0.6822	0.8093	0.7697	0.7906	0.7683
EQM	0.5483	0.3392	0.6335	0.3598	0.7517	0.3864	0.7851	0.3993	1.2496	1.1635	1.5535	1.2435	1.9155	1.5325	2.0822	1.6950

Estimativas de  $\lambda_{11} = 1$

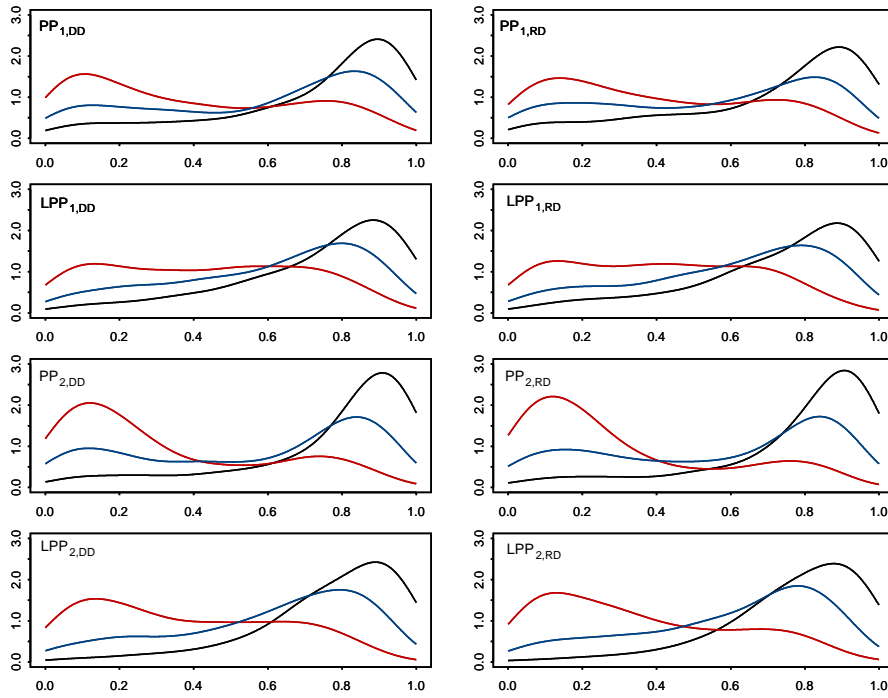
	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>	PP <sub>1,DD</sub>	PP <sub>2,DD</sub>	PP <sub>1,RD</sub>	PP <sub>2,RD</sub>	LPP <sub>1,DD</sub>	LPP <sub>2,DD</sub>	LPP <sub>1,RD</sub>	LPP <sub>2,RD</sub>
	$C_0$															
Média	1.2151	1.0972	1.2433	1.1100	1.1784	1.0658	1.2114	1.0730								
DP	0.3404	0.2301	0.3222	0.2394	0.2949	0.2163	0.3088	0.2312								
EQM	0.1622	0.0624	0.1631	0.0694	0.1188	0.0511	0.1401	0.0588								
	$C_{2,0.05}$								$C_{2,0.10}$							
Média	1.5059	1.4643	1.5677	1.4753	1.5038	1.4564	1.5467	1.4885	1.9395	2.0091	2.0311	2.0882	1.9841	2.0599	2.0531	2.1181
DP	0.4230	0.3589	0.4471	0.3663	0.4475	0.3748	0.4324	0.3836	0.5211	0.5488	0.5730	0.5499	0.5579	0.5650	0.5941	0.5658
EQM	0.4351	0.3446	0.5225	0.3603	0.4543	0.3490	0.4862	0.3860	1.1551	1.3205	1.3927	1.4877	1.2806	1.4437	1.4631	1.5716

**Tabela 5.9:** Estimação dos valores próprios de  $\Sigma_1$ , Modelo 3.

distinguir. Como consequência, as estimativas “projection–pursuit” produzem um enviesamento sistemático para os dados normais. Por outro lado, como  $N_{\text{in}}(1, 2)$  e  $N_{\text{in}}(3, 4)$  são negativos o procedimento “projection–pursuit” baseado em  $f(t) = \ln(t)$  não é capaz de distinguir bem entre  $\beta_1$  e  $\beta_2$  e entre  $\beta_3$  e  $\beta_4$ , conduzindo aos enviesamentos observados nas Tabelas 5.1 e 5.2 e na Figura 5.1. Refira-se ainda que, com a contaminação assimétrica  $C_{2,\varepsilon}$ , estas estimativas têm um comportamento ligeiramente melhor que as baseadas em  $f(t) = t$ .

Por outro lado, no que se refere aos Modelos 2 e 3 todas as populações têm as direcções comuns relacionadas com a ordem decrescente dos valores próprios e como  $\Sigma_{\text{id}}$  têm valores próprios diferentes e bem separados e  $N_{\text{in}}(j, m) > 0, \forall j, m$ , as estimativas de “projection–pursuit” apresentam um desempenho bastante melhor. Esta estrutura ajuda à identificação das direcções comuns, como se pode observar nos grá-



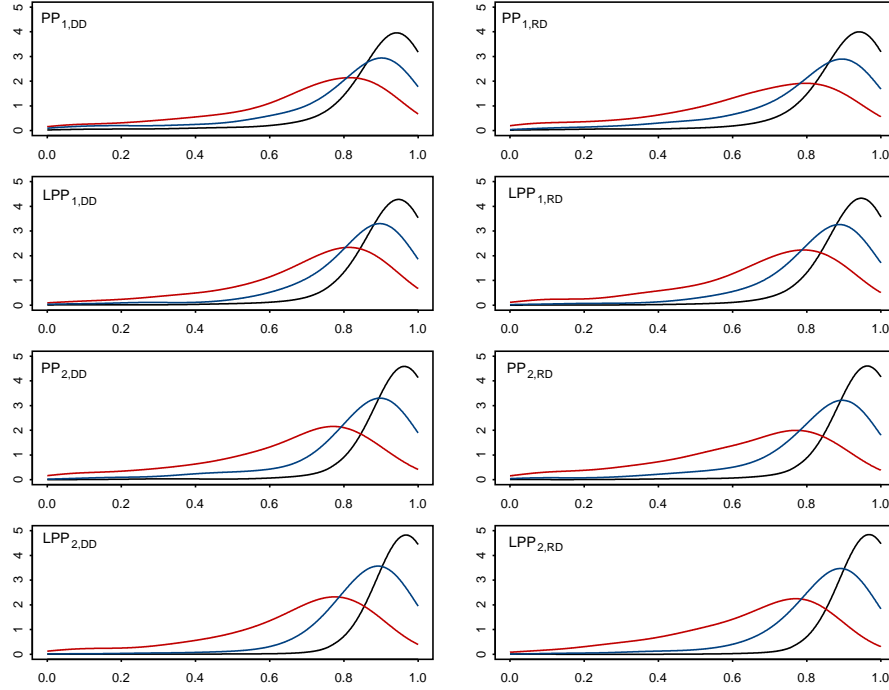


**Figura 5.1:** Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 1. As linhas a preto, azul e vermelho correspondem a  $C_0$ ,  $C_{2,0.05}$  e  $C_{2,0.10}$ , respectivamente.

ficos apresentados. Conforme era “esperado”<sup>1</sup>, no Modelo 2 as estimativas baseadas em  $f(t) = \ln(t)$  têm um desempenho melhor que as construídas com o usual procedimento de “projection-pursuit”, comportamento este que já tinha sido observado na Tabela 5.4. É ainda de salientar que, embora o Modelo 3 não verifique o modelo proporcional, é possível concluir numericamente que  $\alpha = 0$  minimiza (5.53). Não é portanto surpreendente que as estimativas baseadas em  $f(t) = \ln(t)$  se comportem quase sempre melhor que as calculadas com  $f(t) = t$ , como se pode confirmar na Tabela 5.6.

Como já mencionámos, não houve melhoria na estimação dos vectores próprios com a nova técnica de procura de direcções, a técnica RD, em relação à pesquisa nas direcções dos dados normalizados (DD).

<sup>1</sup>O que seria de facto esperado era observar um melhor desempenho com o procedimento PP construído com a função logaritmo em condições assintóticas. No caso finito, o comportamento dos estimadores pode não ser totalmente concordante e daí a colocação das aspas.



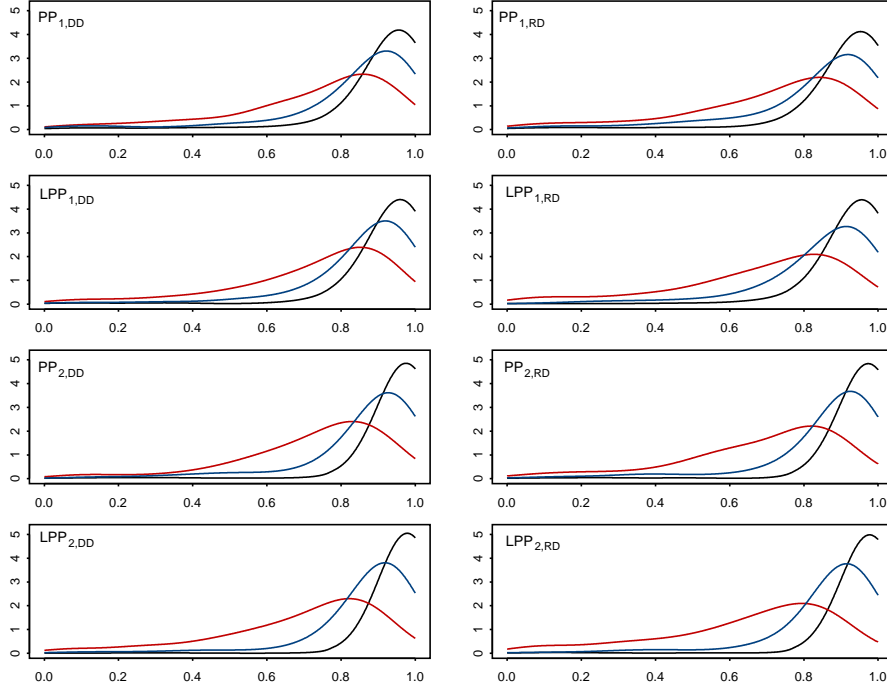
**Figura 5.2:** Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 2. As linhas a preto, azul e vermelho correspondem a  $C_0$ ,  $C_{2,0.05}$  e  $C_{2,0.10}$ , respectivamente.

### 5.6.2 Resultados da simulação em dimensão 2

Foi também desenvolvido uma simulação, sob a validade do modelo CPC, em dimensão  $p = 2$  com  $n_1 = n_2 = n = 50$  e 500 réplicas para avaliar o comportamento dos novos estimadores “projection-pursuit” em maior detalhe. A escolha da dimensão 2 justifica-se porque num espaço bidimensional é possível calcular os estimadores “projection-pursuit” sem recorrer a técnicas de reamostragem. A alternativa para a pesquisa das direcções de projecção utiliza uma grelha construída com direcções fixas e igualmente espaçadas. Na grelha empregue nesta simulação intervêm 1000 direcções. Novamente, as estimativas das direcções comuns foram ordenadas de acordo com a ordenação decrescente dos valores próprios da primeira população.

Uma vez que as estimativas de “projection-pursuit” não são vantajosas quando  $\Sigma_{id}$  têm valores próprios iguais e  $\mathbf{N}_{in}$  entradas negativas, as matrizes escolhidas foram  $\Sigma_1 = \text{diag}(14, 4)$  e  $\Sigma_2 = \text{diag}(12, 2)$ . Neste cenário, pela condição (i) do Teorema 5.1 os estimadores PP gerais são Fisher-consistentes para  $f(t) = t$  e  $f(t) = \ln(t)$ .

Seguindo o trabalho de Boente e Orellana (2001) gerámos amostras normalmente distribuídas e contaminadas, que denotamos, respectivamente, por  $C_0$  e  $C_1$  nas



**Figura 5.3:** Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para o Modelo 3. As linhas a preto, azul e vermelho correspondem a  $C_0$ ,  $C_{2,0.05}$  e  $C_{2,0.10}$ , respectivamente.

Tabelas 5.10 a 5.12 e na Figura 5.4. As amostras contaminadas correspondem a  $\mathbf{X}_{1j}, j = 1, \dots, n$ , independentes e identicamente distribuídas com  $0.80N(\mathbf{0}_2, \mathbf{\Sigma}_1) + 0.1\Delta\boldsymbol{\mu} + 0.1\Delta_{-\boldsymbol{\mu}}$ , com  $\boldsymbol{\mu} = (-0.042, 4.72)^T$ , enquanto que  $\mathbf{X}_{2j}, j = 1, \dots, p$ , são i.i.d. com a distribuição não contaminada  $N(\mathbf{0}_2, \mathbf{\Sigma}_2)$ .

	Mediana				Média			
	PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>	PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>
$C_0$	0.0288	0.0082	0.0266	0.008	0.0643	0.0272	0.0612	0.0221
$C_1$	0.4710	0.4255	0.3746	0.3371	0.4407	0.4946	0.3920	0.4732

**Tabela 5.10:** Mediana e média de  $\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|^2$ .

Na Tabela 5.10 verifica-se que a média e a mediana de  $\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|^2$  são consideravelmente diferentes, consequência do facto de em algumas simulações se atingir grandes valores para esta quantidade. Daí que se tenha resolvido apresentar também a Tabela 5.11, indicando o número de vezes que o valor absoluto do ângulo entre  $\boldsymbol{\beta}_1$  e  $\hat{\boldsymbol{\beta}}_1$  é maior que 45 ( $N_{45}$ ), 60 ( $N_{60}$ ) e 80 ( $N_{80}$ ) graus.

As estimativas da densidade do  $\cos(\hat{\theta})$ , encontram-se na Figura 5.4. Esta figura foi

	PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>	PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>
	$C_0$				$C_1$			
$N_{45}$	6	2	1	1	159	186	117	178
$N_{60}$	3	2	1	1	18	50	20	53
$N_{80}$	1	1	1	0	0	3	1	1

**Tabela 5.11:** Número de vezes,  $N_\alpha$ , que o valor absoluto do ângulo entre  $\beta_1$  e  $\hat{\beta}_1$  é maior que  $\alpha$  graus.

$\lambda$		PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>	PP <sub>1</sub>	PP <sub>2</sub>	LPP <sub>1</sub>	LPP <sub>2</sub>
		$C_0$				$C_1$			
	Média	15.6486	14.1953	15.4463	14.1537	17.4352	12.9794	16.2229	12.6162
14	DP	4.9554	3.8577	4.7895	3.8330	4.5477	2.5372	4.0371	2.1953
	EQM	27.2792	14.9198	25.0354	14.7159	32.5060	7.4812	21.2490	6.7380
		$C_0$				$C_1$			
	Média	4.5133	4.1728	4.4929	4.1764	10.6274	9.6756	10.5011	9.5814
4	DP	1.6508	1.2078	1.5873	1.2022	4.0095	2.7071	4.0699	2.6354
	EQM	2.9893	1.4886	2.7632	1.4765	60.0857	39.6057	58.9132	38.1603

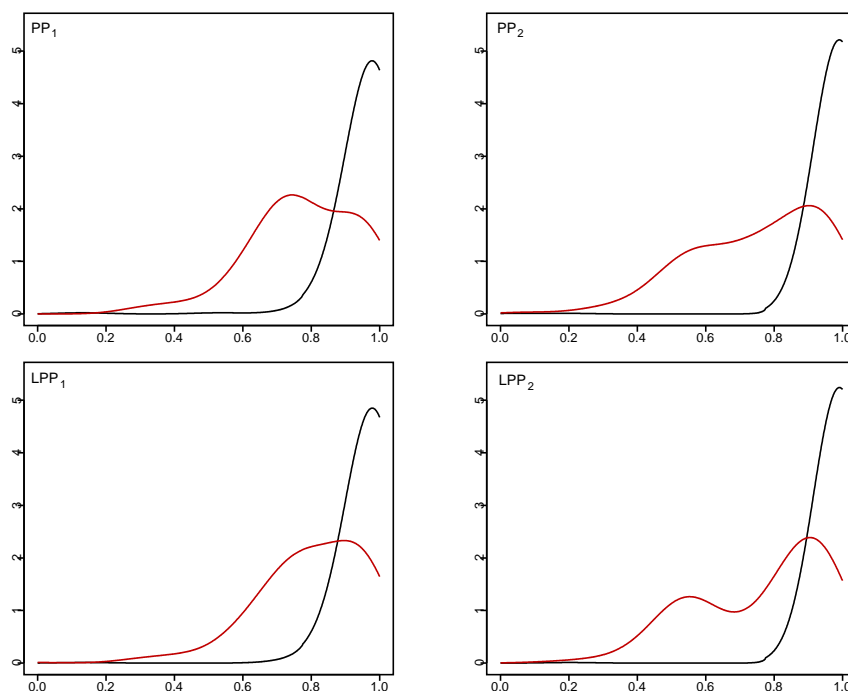
**Tabela 5.12:** Estimação dos valores próprios de  $\Sigma_1$  em dimensão 2.

construída no mesmo estilo da Figura 5.1. A Figura 5.4 e as Tabelas 5.10 e 5.11 mostram também que o método “projection–pursuit” baseado em  $f(t) = \ln(t)$  é mais resistente em especial quando a escala é estimada com o MAD.

Na Tabela 5.12 encontra-se um resumo das medidas sumárias relativas à estimação dos valores próprios de  $\Sigma_1$ . Como se constata, as estimativas “projection–pursuit” baseadas em  $f(t) = \ln(t)$  e que recorrem ao estimador–M de escala apresentam o menor erro quadrático médio, quer sob  $C_0$  quer sob  $C_1$ .

## 5.7 Exemplos com dados reais

Para terminar este capítulo apresentamos os resultados relativos à estimação dos parâmetros do modelo das CPC em dois exemplos reais. A estimação é feita nas duas vertentes abordadas neste trabalho, técnica de “plug–in” e técnica “projection–pursuit”. O desempenho dos novos estimadores PP é avaliado nos dois exemplos que se seguem, o primeiro bidimensional e o segundo em dimensão 5, seleccionados por serem diferentes dos estudados na literatura das *componentes principais comuns*.

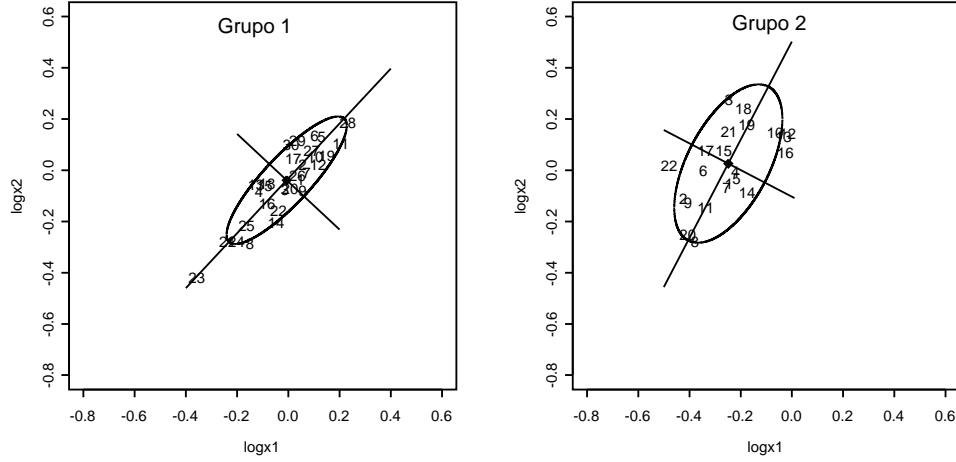


**Figura 5.4:** Estimativa da densidade do coseno do ângulo entre a verdadeira direcção relacionada com o mais pequeno valor próprio da primeira população e a direcção estimada para  $p = 2$ . As linhas a preto e vermelho correspondem a  $C_0$  e  $C_1$ , respectivamente.

### Exemplo 1 (Hemofilia)

Para comprovar o “bom” desempenho dos novos estimadores de “projection–pursuit” em dados reais bidimensionais, considerámos novamente os dados de Hermans e Habbema (1975) referentes a mulheres portadoras e não portadoras do gene da hemofilia. O problema da adequação do modelo CPC clássico a estas observações foi afluído na Subsecção 3.3 do Capítulo 2, onde também apresentámos as estimativas clássicas de PI dos parâmetros do modelo em estudo. O nosso estudo prosseguiu com a estimação robusta dos parâmetros de interesse, embora a comutabilidade das matrizes de dispersão robustas não seja tão evidente como é para o par clássico. A constatação anterior é corroborada pela Figura 5.5 onde se pode ver que os eixos das duas elipses robustas, calculadas com  $(\mathbf{x} - \hat{\boldsymbol{\mu}}_i)^T \mathbf{V}_i (\mathbf{x} - \hat{\boldsymbol{\mu}}_i) = 4$  onde  $\mathbf{V}_i$  e  $\hat{\boldsymbol{\mu}}$  correspondem às estimativas RMCD calculadas com  $h_i$  igual à parte inteira de  $0.75n_i$ , são menos paralelos que os da Figura 2.1. Confrontando as duas figuras, ressalta a ligeira rotação no sentido anti-horário sofrida na elipse do *Grupo 2*. Isto poderá ser justificado pelo uso do estimador RMCD, que neste caso foi calculado considerando as 16 observações “menos” discordantes, i.e., as que constroem a matriz de covariâncias com menor determinante. Como se nota, o contorno de densidade constante para as mulheres portadoras do gene de hemofilia é mais adaptado às 16

observações mais centrais.



**Figura 5.5:** Elipses correspondentes a distância de Mahalanobis robusta constante e eixos das CP robustas.

Na Tabela 5.13 apresentam-se os resultados da estimação dos parâmetros do modelo CPC com o “plug-in” (PI) das matrizes de dispersão estimadas com o RMCD, enquanto que as Tabelas 5.14 e 5.15 reportam os resultados relativos à estimação por “projection-pursuit” construídos com  $f(t) = t$  e  $f(t) = \ln(t)$  e com a escala estimada via MAD e estimador-M.

Embora o bom ajustamento do modelo CPC às matrizes de dispersão RMCD não seja assim tão evidente, os resultados da estimação PI robusta são bastante semelhantes aos obtidos com a análise clássica (ver Tabela 2.3). Esta concordância entre as estimativas clássicas e robustas é justificada pela ausência de observações discordantes na amostra.

Quanto à estimação com a técnica de “projection-pursuit”, a grelha empregue para pesquisar as direcções de projecção foi mais densa que a do estudo de simulação. Considerou-se agora uma malha com 5000 pontos, porque a projecção das observações em apenas 1000 direcções conduzia às mesmas direcções comuns com as duas alternativas em estudo. Isto é consequência da equivalência entre os dois procedimentos de “projection-pursuit”, com  $f(t) = t$  e  $f(t) = \ln(t)$ , quando a estrutura de dispersão das populações é bem modelada por (2.3). Por esta razão, a precisão nos valores apresentados nas Tabelas 5.14 e 5.15 é maior que a considerada na Tabela 5.13.

vectores próprios			valores próprios		
$\log_{10} x_1$	0.658	0.753	<i>Grupo 1</i>	0.027	0.002
$\log_{10} x_2$	0.753	-0.658	<i>Grupo 2</i>	0.027	0.008

**Tabela 5.13:** Estimativas robustas de PI das CPC.

	valores próprios por PP <sub>1</sub>		valores próprios por LPP <sub>1</sub>	
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>
<i>Grupo 1</i>	0.0360	0.0036	0.0363	0.0040
<i>Grupo 2</i>	0.0701	0.0141	0.0697	0.0143

	vectores próprios por PP <sub>1</sub>		vectors próprios por LPP <sub>1</sub>	
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>
$\log_{10} x_1$	0.5755	0.8178	0.5688	0.8225
$\log_{10} x_2$	0.8178	-0.5755	0.8225	-0.5688

**Tabela 5.14:** Estimativas de PP das CPC, com escala estimada através do MAD.

	valores próprios por PP <sub>2</sub>		valores próprios por LPP <sub>2</sub>	
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>
<i>Grupo 1</i>	0.0340	0.0027	0.0341	0.0027
<i>Grupo 2</i>	0.0482	0.0109	0.0482	0.0109

	vectors próprios por PP <sub>2</sub>		vectors próprios por LPP <sub>2</sub>	
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>
$\log_{10} x_1$	0.6518	0.7584	0.6504	0.7596
$\log_{10} x_2$	0.7584	-0.6518	0.7596	-0.6504

**Tabela 5.15:** Estimativas de PP das CPC, com escala estimada através do estimador-M.

É notório uma grande concordância entre as estimativas PP<sub>2</sub> e LPP<sub>2</sub> e as construídas via “plug-in”.

### Exemplo 2 (Variedades de Castanheiros)

Como segundo exemplo considerámos um conjunto de dados com 5 variáveis e 2 grupos. Estes dados são parte dos utilizados por Oliveira (1995) onde foi aplicada uma análise em *componentes principais* e referem-se a características morfométricas de folhas de castanheiro medidas em duas variedades (*Lada* e *Longal*, com 100 e 47 observações, respectivamente), ano de 1989, através das variáveis:<sup>2</sup>

$x_1$ : comprimento do pecíolo (em *mm*);

$x_2$ : número de nervuras do lado direito da folha;

<sup>2</sup>Ver Apêndice B.

$x_3$ : número de nervuras do lado esquerdo da folha;

$x_4$ : número de dentes do lado direito da folha;

$x_5$ : número de dentes do lado esquerdo da folha.

O estudo foi iniciado com uma análise em *componentes principais* robusta para cada variedade em separado, recorrendo ao estimador RMCD e fixando  $h_i = [0.75n_i]$ . Os valores da Tabela 5.16 mostram que as duas variedades apresentam eixos principais similares mas com variabilidades distintas, corroborando a validade do modelo CPC. Por outro lado, esta conclusão já não é tão evidente nos resultados das *componentes principais* clássicas, apresentados na Tabela 5.17, o que poderá, eventualmente, ser prenúncio da existência de observações discordantes. As estimativas robustas das

	<i>Lada</i>					<i>Longal</i>				
	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>
valores próprios	26.962	7.386	1.985	1.918	0.624	66.208	19.581	2.670	0.603	0.465
	vectores próprios					vectores próprios				
$x_1$	-0.622	0.781	-0.015	-0.061	0.006	-0.757	0.649	0.071	0.010	0.001
$x_2$	0.492	0.383	-0.645	0.002	-0.441	0.390	0.508	-0.386	-0.562	-0.353
$x_3$	0.423	0.389	0.257	0.642	0.438	0.376	0.377	0.573	-0.168	0.600
$x_4$	0.339	0.209	-0.000	-0.730	0.556	0.280	0.368	-0.481	0.701	0.250
$x_5$	0.278	0.221	0.719	-0.229	-0.552	0.232	0.208	0.535	0.405	-0.673

**Tabela 5.16:** Análise robusta em CP.

	<i>Lada</i>					<i>Longal</i>				
	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>
valores próprios	27.023	11.191	2.314	1.991	0.872	59.241	43.088	3.572	1.499	0.547
	vectores próprios					vectores próprios				
$x_1$	-0.565	0.823	-0.026	0.043	-0.024	-0.920	0.386	-0.021	-0.055	0.034
$x_2$	0.531	0.346	-0.466	0.348	0.509	0.227	0.512	-0.504	-0.349	-0.557
$x_3$	0.460	0.324	-0.245	-0.600	-0.513	0.244	0.485	0.412	-0.529	0.505
$x_4$	0.339	0.199	0.443	0.624	-0.510	0.166	0.407	-0.470	0.571	0.510
$x_5$	0.269	0.240	0.725	-0.357	0.465	0.122	0.433	0.595	0.519	-0.417

**Tabela 5.17:** Análise clássica em CP.

*componentes principais comuns* obtidas com o “plug-in” das matrizes de dispersão estimadas com o RMCD, considerando  $h_i = [0.75n_i]$ , são apresentadas na Tabela 5.18 em conjunto com as estimativas clássicas. As Tabelas 5.19, 5.20, 5.21 e 5.22 reportam as estimativas de “projection-pursuit”, para os procedimentos com  $f(t) = t$  e  $f(t) = \ln(t)$ , baseados no MAD e no estimador-M e nas técnicas DD e RD para pesquisa de direcções.



	valores próprios por PI (RMCD)					valores próprios por PI (clássico)				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
<i>Lada</i>	26.904	7.392	1.993	1.949	0.637	26.968	11.138	2.047	2.362	0.876
<i>Longal</i>	65.318	20.391	2.734	0.588	0.496	54.673	47.508	3.665	1.548	0.552

	vectores próprios por PI (RMCD)					vectores próprios por PI (clássico)				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
$x_1$	0.661	0.750	0.022	0.012	-0.002	0.579	0.812	-0.039	-0.051	-0.032
$x_2$	-0.463	0.429	-0.444	-0.409	0.488	-0.497	0.329	-0.492	-0.348	0.530
$x_3$	-0.417	0.355	0.585	-0.359	-0.478	-0.454	0.293	0.442	-0.493	-0.520
$x_4$	-0.325	0.290	-0.469	0.576	-0.508	-0.355	0.250	-0.441	0.607	-0.499
$x_5$	-0.262	0.208	0.491	0.610	0.525	-0.295	0.288	0.606	0.515	0.445

**Tabela 5.18:** Estimativas de PI das CPC.

	valores próprios por PP <sub>1,DD</sub>					valores próprios por LPP <sub>1,DD</sub>				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
<i>Lada</i>	33.405	12.673	2.584	2.357	2.066	33.405	12.673	3.156	2.571	1.930
<i>Longal</i>	102.388	29.096	7.969	3.844	1.145	102.388	29.097	5.743	2.752	2.671

	vectores próprios por PP <sub>1,DD</sub>					vectores próprios por LPP <sub>1,DD</sub>				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
$x_1$	0.903	0.269	0.332	0.033	-0.019	0.904	0.269	-0.263	-0.061	0.197
$x_2$	-0.177	0.701	-0.012	-0.686	0.083	-0.177	0.701	0.124	0.643	0.221
$x_3$	-0.189	0.453	0.064	0.418	-0.762	-0.190	0.453	0.273	-0.731	0.388
$x_4$	-0.341	0.114	0.833	0.233	0.351	-0.341	0.114	-0.914	-0.097	0.160
$x_5$	0.013	0.468	-0.438	0.547	0.538	0.013	0.468	-0.076	-0.197	-0.858

**Tabela 5.19:** Estimativas de PP das CPC, com escala estimada através do MAD e técnica “data direction”.

	valores próprios por PP <sub>2,DD</sub>					valores próprios por LPP <sub>2,DD</sub>				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
<i>Lada</i>	32.547	10.358	2.416	2.351	0.966	33.849	7.375	2.472	2.061	0.764
<i>Longal</i>	88.693	21.916	3.307	1.832	1.019	85.613	21.085	3.820	1.130	0.986

	vectores próprios por PP <sub>2,DD</sub>					vectores próprios por LPP <sub>2,DD</sub>				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
$x_1$	0.886	0.451	0.085	-0.061	-0.027	-0.797	0.594	0.104	-0.007	0.042
$x_2$	-0.299	0.612	0.344	0.457	0.458	0.338	0.578	-0.582	0.378	-0.266
$x_3$	-0.289	0.381	0.385	-0.754	-0.235	0.346	0.328	0.660	0.457	0.359
$x_4$	-0.182	0.425	-0.366	0.328	-0.739	0.291	0.372	-0.152	-0.686	0.533
$x_5$	-0.095	0.311	-0.769	-0.335	0.435	0.218	0.261	0.439	-0.422	-0.717

**Tabela 5.20:** Estimativas de PP das CPC, com escala estimada através do estimador–M e técnica “data direction”.

Conforme era esperado, observam-se diferenças entre as estimativas clássicas e robustas, mas mais significativas para os valores próprios do segundo grupo. Além disso, é notória a diferença entre as estimativas obtidas pelas técnicas de “plug-in” e “projection-pursuit” provavelmente consequência da presença de “inliers” nas observações projectadas. É interessante ainda referir que, de uma maneira geral, a primeira direcção comum estimada pelo método PP usual e pela nova proposta com

	valores próprios por $PP_{1, RD}$					valores próprios por $LPP_{1, RD}$				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
<i>Lada</i>	35.033	11.953	4.393	2.210	0.906	35.043	13.634	3.388	2.316	1.257
<i>Longal</i>	99.510	32.257	7.769	6.131	1.047	99.471	25.617	7.629	3.071	1.067

	vectores próprios por $PP_{1, RD}$					vectores próprios por $LPP_{1, RD}$				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
$x_1$	-0.934	0.301	-0.004	-0.187	0.039	-0.934	0.250	0.234	0.084	0.055
$x_2$	0.202	0.788	0.261	0.148	-0.498	0.203	0.641	0.014	0.596	-0.440
$x_3$	0.273	0.362	0.164	-0.664	0.572	0.273	0.144	0.745	0.157	0.571
$x_4$	0.045	0.393	-0.558	0.529	0.504	0.045	0.602	-0.514	-0.235	0.563
$x_5$	0.100	0.058	-0.771	-0.472	-0.412	0.099	0.380	0.355	-0.747	-0.401

**Tabela 5.21:** Estimativas de PP das CPC, com escala estimada através do MAD e técnica “random direction”.

	valores próprios por $PP_{2, RD}$					valores próprios por $LPP_{2, RD}$				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
<i>Lada</i>	32.973	7.674	2.270	2.177	1.401	32.975	7.674	2.412	2.057	1.410
<i>Longal</i>	84.698	20.388	3.863	2.308	1.156	84.704	20.388	3.520	2.163	1.138

	vectores próprios por $PP_{2, RD}$					vectores próprios por $LPP_{2, RD}$				
	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>	CPC <sub>1</sub>	CPC <sub>2</sub>	CPC <sub>3</sub>	CPC <sub>4</sub>	CPC <sub>5</sub>
$x_1$	-0.814	0.570	0.085	0.070	0.006	-0.814	0.570	0.099	0.048	0.008
$x_2$	0.414	0.532	-0.087	0.626	-0.382	0.414	0.532	0.033	0.555	0.486
$x_3$	0.300	0.349	0.449	0.033	0.766	0.300	0.349	0.492	0.058	-0.737
$x_4$	0.147	0.372	-0.800	-0.366	0.258	0.147	0.374	-0.844	-0.120	-0.337
$x_5$	0.232	0.363	0.380	-0.684	-0.449	0.232	0.363	0.188	-0.820	0.327

**Tabela 5.22:** Estimativas de PP das CPC, com escala estimada através do estimador–M e técnica “random direction”.

a aplicação da função logaritmo são semelhantes.

Com os exemplos apresentados e o estudo de simulação desenvolvido pensamos que ficou evidente a vantagem da estimação robusta das *componentes principais comuns* e, também, do emprego da nova técnica de “projection–pursuit”. Como se viu, a estimação robusta fornece bastantes vantagens em relação à estimação na vertente clássica. Enquanto que para dados “limpos” ou não contaminados os resultados dos estimadores robustos são semelhantes aos clássicos, para os dados “contaminados” os seus resultados são francamente melhores.



## Capítulo 6

# Detecção de Observações Discordantes/Influentes

### 6.1 Introdução

Embora o problema de detecção e tratamento de observações discordantes (“outliers”) venha sendo considerado desde os primórdios da estatística, tem recebido nas últimas décadas um especial interesse. Isto deve-se principalmente às possibilidades de análise de conjuntos de dados cada vez mais volumosos e complexos, proporcionadas pelos modernos meios computacionais.

Como vimos no Capítulo 3, Secção 3, este tópico tem mesmo uma forte interrelacção com a estatística robusta. Daí que, em muitos dos trabalhos estatísticos que enveredam por esta abordagem lhe seja dado um destaque especial. Actualmente, a literatura existente sobre “outliers” é bastante extensa, havendo mesmo alguns livros que apresentam uma análise aprofundada deste tópico. Podemos citar, entre outros, os livros “Outliers in Statistical Data” de Barnett e Lewis (1978, 1984, 1994), “Identification of Outliers” de Hawkins (1980) e “Robust Regression and Outlier Detection” dos autores Rousseeuw e Leroy (1987). Nos trabalhos existentes têm sido sugeridos vários métodos para a detecção de “outliers”, tanto para o caso univariado como para o multivariado. Para esta última situação, o problema não é nada trivial especialmente quando existem várias observações discordantes. Quando se trabalha com pontos num espaço de dimensão  $p$  superior a 2 ou 3 as técnicas gráficas deixam de funcionar como análise exploratória de eventuais observações discordantes. Além disso, quando  $p$  é elevado torna-se difícil construir e analisar todos os  $\frac{p(p-1)}{2}$  gráficos bidimensionais. Por outro lado, é importante não esquecer que as observações

multivariadas podem até parecer concordantes quando é feita a inspecção de cada variável isoladamente ou mesmo dos  $\frac{p(p-1)}{2}$  pares de variáveis. No entanto, isto não significa que a localização no espaço  $p$  dimensional esteja próxima da nuvem descrita pela maioria das observações. Como ilustração deste tipo de “outliers”, veja-se o exemplo apresentado na página 7 do livro de Rousseeuw e Leroy (1987). Por outro lado, a tentativa da identificação de “outliers” multivariados com outros métodos, não robustos, não é frequentemente uma boa alternativa. A maioria das propostas clássicas utiliza o valor da distância de Mahalanobis (1936) como um indicador da possível discordância da observação. Esta técnica, originalmente introduzida por Healy (1968), pode facilmente falhar, visto que tanto a média como a matriz de covariâncias amostral, ambas com um papel fundamental no cálculo da distância de Mahalanobis, podem ser afectadas pelas observações discordantes. Daí que, quando há “outliers” múltiplos, alguns deles não apresentem um valor elevado da distância de Mahalanobis. Este efeito, conhecido na literatura por “masking effect”, não deve ser esquecido e é importante que os procedimentos utilizados na detecção de observações discordantes o evitem.

Como já referimos, existe uma forte interligação entre a detecção de “outliers” e a estimação robusta. Alguns autores abordaram estes dois tópicos em simultâneo, tendo sido o trabalho de Gnanadesikan e Kettenring (1972) pioneiro nesta perspectiva. Para tentar evitar a “má” influência dos “outliers”, os autores anteriormente referidos propõem a exclusão de todas as observações mais distanciadas (segundo a distância de Mahalanobis). Após esta operação é calculada uma nova média e matriz de covariâncias amostral e assim sucessivamente.

Outra questão que também preocupa os estatísticos, em especial os mais aplicados, é a identificação das observações que podem afectar o desempenho dos estimadores. Um “outlier”, embora seja discordante do “miolo” dos dados, consoante a sua localização espacial, pode ter ou não uma grande influência no resultado da estimação. Em consequência, a sua remoção pode afectar substancialmente, ou não, os resultados da análise. Por outro lado, as observações cuja extracção provoca uma alteração significativa na análise estatística são referidas como influentes. A maioria das observações influentes mostra-se como “outliers”, segundo algum critério, mas nem todos os “outliers” são observações influentes. Além disso, a classificação de uma observação como influente está dependente da análise efectuada. Para uma determinada análise pode ainda acontecer que a mesma observação exerça uma influência distinta nos vários parâmetros de interesse.

Em conclusão, é extremamente importante dispor de métodos para detectar si-

multaneamente observações discordantes e/ou influentes e que evitem o indesejável “masking effect”.

No que se refere ao modelo das *componentes principais* foram vários os autores que abordaram este problema. Podemos citar, entre outros, Gnanadesikan e Kettenring (1972), Devlin *et al.* (1981), Critchley (1985), Pack *et al.* (1988), Tanaka (1988), Pires (1990), Shi (1997) e Croux e Haesbroeck (1999b) como proponentes de técnicas de diagnóstico para a análise em *componentes principais*. Alguns destes autores, como referimos no Capítulo 4, construíram métodos para a detecção de observações discordantes/influentes que recorrem à *função de influência* dos funcionais valores e vectores próprios. O próprio Hampel (1974) já tinha indicado a *função de influência* como uma ferramenta apropriada para estudar simultaneamente problemas de detecção de “outliers” e de estimação robusta. No entanto, como a verdadeira distribuição e o verdadeiro valor das características em estudo são desconhecidos considera-se geralmente uma versão empírica da *função de influência*. Existem várias definições para a *função de influência empírica*. Algumas delas foram apresentadas no Capítulo 3, Subsecção 5.4. Critchley (1985), por exemplo, propõe o emprego da *função de influência empírica* (EIF) criada tomando na *função de influência* uma estimativa,  $\hat{F}$ , da função de distribuição  $F$ . Para os valores e vectores próprios da matriz de covariâncias amostral  $\mathbf{S}$ , resulta desta abordagem que

$$\text{EIF}(\mathbf{x}_i, \lambda_{\mathbf{S},j}) = \text{IF}\left(\mathbf{x}_i, \lambda_{\mathbf{S},j}, \hat{F}\right) = y_{ij}^2 - \hat{\lambda}_j \quad (6.1)$$

e

$$\text{EIF}(\mathbf{x}_i, \boldsymbol{\beta}_{\mathbf{S},j}) = \text{IF}\left(\mathbf{x}_i, \boldsymbol{\beta}_{\mathbf{S},j}, \hat{F}\right) = y_{ij} \sum_{k \neq j} \frac{y_{ik} \hat{\boldsymbol{\beta}}_k}{\hat{\lambda}_j - \hat{\lambda}_k}, \quad (6.2)$$

com  $y_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})^T \hat{\boldsymbol{\beta}}_j$ . Como se observa, a proposta deste autor apoia-se na matriz de covariâncias amostral e além disso utiliza estimadores clássicos para estimar os parâmetros desconhecidos. O autor Shi (1997) também considera este estimador da matriz de covariâncias. Por outro lado, Jaupi e Saporta (1993) optam pela estimação robusta da matriz de dispersão, empregando um estimador-M, com a introdução de novas medidas empíricas robustas para quantificar a influência de cada observação. Mais recentemente, os autores Croux e Haesbroeck (1999b) confrontam a utilização da matriz de covariâncias amostral da proposta de Critchley (1985) com outra proposta baseada no estimador RMCD (Rousseeuw, 1985). Croux e Haesbroeck (1999b) salientam a fraca influência das observações contaminadas na EIF dos funcionais valores e vectores próprios, quando se estima a matriz de dispersão com o RMCD. Referem também que o efeito das observações influentes na EIF clássica, construída com o estimador clássico da matriz de covariâncias, pode ser surpreen-

dente. Neste caso pode acontecer, por exemplo, que uma observação influente origine um pequeno valor da EIF. Segundo estes autores, a observação da EIF poderá ser útil na quantificação do grau de influência das observações mas não é conveniente utilizá-la na detecção de observações discordantes.

Por outro lado, para o modelo das *componentes principais comuns* até à data não foram apresentadas propostas para a detecção de observações discordantes/influente. Daí que tenhamos reservado algum tempo da nossa investigação para abordar esta questão. Além disso, este tópico é mesmo essencial para o modelo em que trabalhamos. Note-se que, no caso do modelo das *componentes principais comuns*, a menos que se pretenda efectuar procedimentos inferenciais sobre os parâmetros de interesse, não estão presentes imposições distribucionais. Isto faz com que neste caso as técnicas robustas devam incidir essencialmente sobre as possíveis observações discordantes/influente, não sendo necessário tanta preocupação com os problemas decorrentes dos desvios distribucionais.

Assim, com base nas *funções de influência parciais* construímos medidas para a detecção de observações discordantes/influente. Estes resultados encontram-se na Secção 2. Para avaliar o desempenho das nossas propostas, na Secção 3 utilizamos as novas medidas na detecção de eventuais observações discordantes/influente nos dados estudados por Hermans e Habbema (1975) e por Oliveira (1995).

## 6.2 Novas propostas

Como se sabe, as observações consideradas como “outliers” normalmente exercem uma pequena influência nos estimadores robustos (resistentes). Esta constatação é corroborada pelas ligeiras variações das EIF de estimadores resistentes da matriz de dispersão. Por outro lado, como Croux e Haesbroeck (1999b) concluem, para o modelo das *componentes principais*, utilizando as EIF com estimadores clássicos o “masking effect” poderá camuflar eventuais observações discordantes. De forma a contornar estas questões Pison *et al.* (2000) propõem, no contexto da *análise factorial*, utilizar a EIF clássica mas considerando estimadores robustos dos parâmetros desconhecidos. Para o caso da análise em *componentes principais*, a ideia destes autores traduz-se em aplicar as equações (6.1) e (6.2) onde  $\hat{\lambda}$  e  $\hat{\beta}$  referem estimadores robustos dos valores e vectores próprios, respectivamente. Segundo Pison *et al.* (2000) este método é bastante eficiente uma vez que os estimadores robustos  $\hat{\lambda}$  e  $\hat{\beta}$  não são afectados pelas observações discordantes e por isso fornecem boas aproximações dos verdadeiros parâmetros  $\lambda$  e  $\beta$ . Além disso, na prática obtêm-se

bons resultados na detecção de observações influentes. Isto é consequência directa da EIF robusta da matriz de covariâncias  $\mathbf{S}$ , i.e., calculada com o “plug-in” dos estimadores  $\hat{\boldsymbol{\lambda}}$  e  $\hat{\boldsymbol{\beta}}$  robustos, evitar o indesejável “masking effect”.

Generalizando esta questão para  $k > 1$  populações, é esperado que os efeitos sentidos nas  $\text{PIF}_i$  dos estimadores clássicos ou robustos sejam semelhantes aos referidos anteriormente. De forma a contornar estes obstáculos, a nossa sugestão é considerar uma versão robusta das  $\text{PIF}_i$  dos estimadores clássicos  $\mathbf{S}_i$ , seguindo a sugestão de Pison *et al.* (2000). Este procedimento é também análogo ao desenvolvido por Rousseeuw e van Zomeren (1990) na construção da versão robusta da distância de Mahalanobis.

No entanto, esta abordagem pode tornar-se problemática em problemas com dimensão elevada, devido ao número elevado de *funções de influência parciais*, uma por cada parâmetro univariado. A análise simplificar-se-ia se fosse possível observar um número mais reduzido de valores. Neste sentido, considerámos somente duas medidas, para cada população, uma para os valores próprios e outra para os vectores próprios comuns. Para colmatar eventuais problemas consequentes das diferentes ordens de grandeza dos valores próprios considerámos versões empíricas das  $\text{PIF}_i$  robustas e standardizadas. Desta forma consegue-se quantificar efectivamente qual a influência que determinada observação exerce na análise das CPC. Por outro lado, no que se refere aos vectores próprios, as medidas de diagnóstico devem ser invariantes para transformações ortogonais. Isto é alcançado transformando o problema para o caso diagonal uma vez que, pela Proposição 4.1,  $\text{PIF}_i\{\mathbf{x}, \boldsymbol{\beta}, F\} = \boldsymbol{\beta} \text{PIF}_i\{\boldsymbol{\beta}^T \mathbf{x}, \boldsymbol{\beta}, F_0\}$ , com  $F_0 = F_{1,0} \times \cdots \times F_{k,0}$ , sendo  $F_{i,0}$  a distribuição de  $\boldsymbol{\beta}^T \mathbf{y}$ ,  $\mathbf{y} \sim F_i$  e  $\boldsymbol{\beta}(F)$  um funcional dos vectores próprios *equivariante*. Estas considerações permitem apresentar a definição que se segue.

**Definição 6.1 (medidas de diagnóstico para os valores e vectores próprios)**

Dada uma observação  $\mathbf{x}$  proveniente da  $i$ -ésima população, considere-se

$$\text{IML}_i(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \left\{ \sum_{r=1}^p \frac{\text{PIF}_i(\mathbf{x}, \lambda_{\mathbf{S},ir}, F)^2}{v_{ir}(\boldsymbol{\beta}, \boldsymbol{\lambda})} \right\}^{\frac{1}{2}}, \quad (6.3)$$

$$\text{IMB}_i(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \left[ \sum_{r=1}^p \left\{ \text{PIF}_i(\boldsymbol{\beta}^T \mathbf{x}, \boldsymbol{\beta}_{\mathbf{S},r}^{(r)}, F_0) \right\}^T \mathbf{A}_{ir}^{-1}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \left\{ \text{PIF}_i(\boldsymbol{\beta}^T \mathbf{x}, \boldsymbol{\beta}_{\mathbf{S},r}^{(r)}, F_0) \right\} \right]^{\frac{1}{2}}, \quad (6.4)$$

onde  $\lambda_{\mathbf{S},il}$  e  $\boldsymbol{\beta}_{\mathbf{S},r}$  denotam os funcionais relacionados com os estimadores clássicos,  $\mathbf{z}^{(r)}$  representa o vector  $\mathbf{z}$  sem a  $r$ -ésima componente,  $\boldsymbol{\lambda} = (\lambda_{11}, \dots, \lambda_{1p}, \dots, \lambda_{k1}, \dots, \lambda_{kp})^T$



e  $\boldsymbol{\beta}$  são os parâmetros desconhecidos e

$$\begin{aligned}\mathbf{A}_{ir}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= E_{F_i} \left\{ \text{PIF}_i(\mathbf{U}, \boldsymbol{\beta}_{\mathbf{S},r}^{(r)}, F_0) \text{PIF}_i(\mathbf{U}, \boldsymbol{\beta}_{\mathbf{S},r}^{(r)}, F_0)^T \right\}, \\ v_{ir}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= E_{F_i} \left\{ \text{PIF}_i(\mathbf{U}, \boldsymbol{\lambda}_{\mathbf{S},ir}, F) \right\}^2.\end{aligned}$$

As medidas para detecção de observações discordantes/influentes definem-se como  $\text{IML}_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$  e  $\text{IMB}_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ , onde  $\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\lambda}}$  denotam a substituição dos parâmetros desconhecidos por estimadores robustos.

A  $r$ -ésima coordenada foi excluída de  $\text{IMB}_i(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ ,  $i = 1, \dots, k$ , porque no caso diagonal tanto as  $k$  *funções de influência parciais* como as  $k$  variâncias são iguais a zero.

Outra alternativa seria utilizarmos as *funções de influência parciais* dos funcionais relacionados com os estimadores de “projection–pursuit” construídos com o estimador clássico variância amostral. Contudo, confrontando (4.36) e (4.37) com (4.68) e (4.69), facilmente se verifica que ambas as expressões são equivalentes.

Por outro lado, quando  $F_{i,0} = N(\mathbf{0}_p, \boldsymbol{\Lambda}_i)$  de (4.36) e (4.37) as expressões de diagnóstico simplificam, ficando

$$\text{IML}_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = \left[ \sum_{r=1}^p \frac{\left\{ \left( \hat{\boldsymbol{\beta}}_r^T \mathbf{x} \right)^2 - \hat{\lambda}_{ir} \right\}^2}{2\hat{\lambda}_{ir}^2} \right]^{\frac{1}{2}}, \quad (6.5)$$

$$\text{IMB}_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = \left[ \sum_{r=1}^p \sum_{s \neq r} \frac{\left\{ \left( \hat{\boldsymbol{\beta}}_r^T \mathbf{x} \right) \left( \hat{\boldsymbol{\beta}}_s^T \mathbf{x} \right) \right\}^2}{\hat{\lambda}_{ir} \hat{\lambda}_{is}} \right]^{\frac{1}{2}}, \quad (6.6)$$

com  $\hat{\boldsymbol{\beta}}$  e  $\hat{\lambda}_{ij}$  referindo estimadores robustos construídos por “plug–in” ou “projection–pursuit”. Note-se que as medidas IML e IMB são funções simples dos “scores” estandardizados, isto é,  $\left( \hat{\boldsymbol{\beta}}_r^T \mathbf{x} / \sqrt{\hat{\lambda}_{ir}} \right)$ . É evidente a vantagem de IML e IMB na análise da influência das observações nos parâmetros de interesse, face à construção de gráficos individuais para cada parâmetro, uma vez que reduz o número de gráficos de  $\left\{ p + \frac{1}{2} p(p-1) \right\} k$  para  $2k$ .

Para a detecção das observações influentes, devemos comparar os valores das medi-

das IML e IMB com percentis elevados das variáveis aleatórias

$$G_{\lambda} = \left\{ \sum_{r=1}^p \frac{(Z_r^2 - 1)^2}{2} \right\}^{\frac{1}{2}},$$

$$G_{\beta} = \left( \sum_{r=1}^p \sum_{s \neq r} Z_r^2 Z_s^2 \right)^{\frac{1}{2}} = \left\{ \left( \sum_{r=1}^p Z_r^2 \right)^2 - \sum_{r=1}^p Z_r^4 \right\}^{\frac{1}{2}},$$

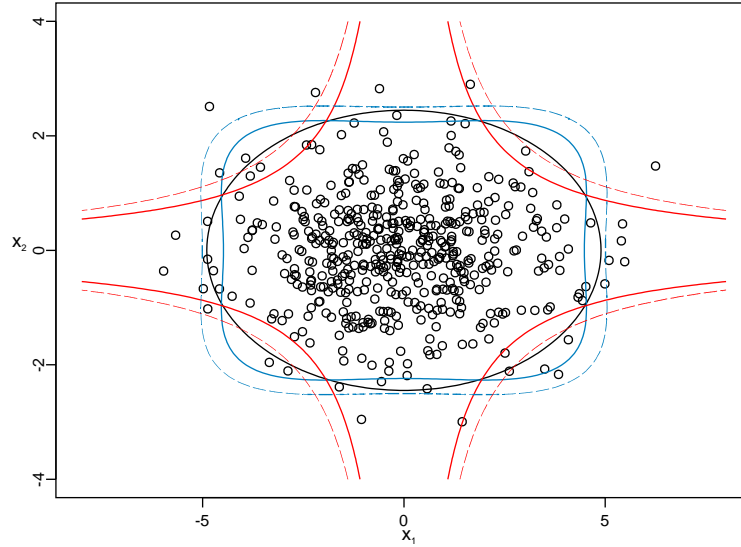
respectivamente, onde  $Z_1, \dots, Z_p$  são i.i.d. com a distribuição  $N(0, 1)$ .

A Tabela 6.1 indica estimativas dos percentis 95%, 97.5% e 99% de  $G_{\lambda}$  e  $G_{\beta}$ , denotados por  $IML_{0.95}$ ,  $IML_{0.975}$ ,  $IML_{0.99}$ ,  $IMB_{0.95}$ ,  $IMB_{0.975}$  e  $IMB_{0.99}$ , respectivamente, e obtidos por simulação. Para cada dimensão  $p$ , gerámos 100 réplicas de amostras de dimensão  $m = 10000$ . Os valores tabelados são as medianas das 100 réplicas dos correspondentes percentis empíricos de  $G_{\lambda,1}, \dots, G_{\lambda,m}$  e  $G_{\beta,1}, \dots, G_{\beta,m}$  com  $m = 10000$ .

$p$	2	3	5	10	$p$	2	3	5	10
$IML_{0.95}$	2.927	3.514	4.316	5.604	$IMB_{0.95}$	3.088	5.058	8.463	15.910
$IML_{0.975}$	3.788	4.381	5.215	6.494	$IMB_{0.975}$	3.933	6.131	9.846	17.190
$IML_{0.99}$	4.974	5.558	6.375	7.637	$IMB_{0.99}$	5.088	7.535	11.630	20.190

**Tabela 6.1:** Estimativas dos percentis 95%, 97.5% e 99% de  $G_{\lambda}$  e  $G_{\beta}$ .

É ainda interessante referir que, uma vez que o modelo das *componentes principais* é um caso particular do modelo CPC com  $k = 1$ , as medidas de diagnóstico aqui propostas também podem ser utilizadas nessa situação. A comparação dos resultados alcançados com estas novas medidas e outras existentes na literatura permitiria avaliar o desempenho desta metodologia na análise em *componentes principais*. Contudo, esta questão está fora do âmbito da nossa investigação embora seja interessante para trabalho futuro. No entanto, para ilustrarmos como funciona a nossa proposta na detecção de observações discordantes/influentes comparámos os desempenhos de IML e IMB com os obtidos através da distância de Mahalanobis. Pareceu-nos relevante esta comparação visto que a distância de Mahalanobis é uma das mais utilizadas na detecção de “outliers” multivariados. Assim, na Figura 6.1 apresentam-se as regiões de detecção de “outliers” usando a distância de Mahalanobis com  $\Sigma = \text{diag}(4, 1)$ ,  $IML_1(\mathbf{x}, \beta, \lambda)$  e  $IMB_1(\mathbf{x}, \beta, \lambda)$  com  $\beta = \mathbf{I}_2$  e  $\lambda = (4, 1)$ , conjuntamente com 500 observações normalmente distribuídas de média  $(0, 0)^T$  e matriz de covariâncias  $\Sigma = \text{diag}(4, 1)$ . As linhas sólidas indicam os contornos de detecção



**Figura 6.1:** Regiões de detecção obtidas com a distância de Mahalanobis (elipse),  $IML_1$  (curvas a azul) e  $IMB_1$  (curvas a vermelho). Os limites de detecção a 95% são indicados com curvas contínuas e os limites a 97.5% com curvas tracejadas.

a 95% para  $IML_1$  e  $IMB_1$  enquanto que as linhas a tracejado referem os limites de detecção a 97.5%. Como é óbvio, a elipse demarca a região construída com a distância de Mahalanobis. Convém realçar que rigorosamente, e usando o método de Bonferroni para comparações múltiplas, deveriam usar-se os percentis 97.5% de  $IML_1$  e  $IMB_1$  e o percentil 95% da distância de Mahalanobis.

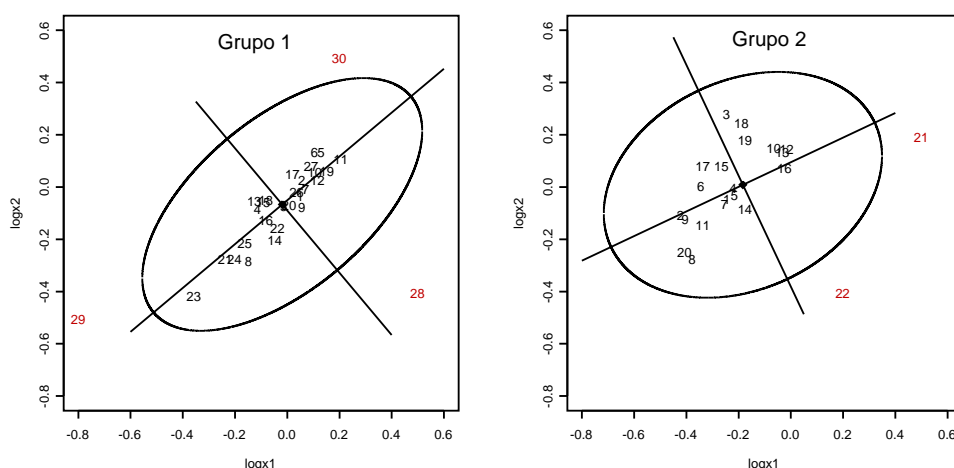
## 6.3 Exemplos

O desempenho das medidas IML e IMB na detecção de eventuais observações discordantes/influentes foi avaliado considerando novamente os dois exemplos tratados no Capítulo 5, Secção 7. Para o cálculo das quantidades IML e IMB desenvolvemos dois programas implementados na linguagem do S-PLUS versão 2000. As listagens dos códigos desses programas encontram-se no Apêndice C.

### Exemplo 1 (Hemofilia)

No Capítulo 2, Subsecção 3.3, apresentámos na Figura 2.1 as elipses correspondentes a distância de Mahalanobis constante e igual a 2, os quais traduzem uma confiança de aproximadamente 90%. Analisando as regiões limitadas por esses dois contornos, conjuntamente com as 30 observações das mulheres não portadoras de hemofilia e as 22 referentes às mulheres portadoras desse gene, não parecem existir observações que se destaquem como discordantes. Eventualmente, a única excepção é a observação

23 do *Grupo 1*. Além disso, aumentando a confiança destes contornos para 95% quanto muito diminuía o número das observações em destaque. Para investigarmos o desempenho das medidas propostas, introduzimos contaminação determinística modificando três observações no *Grupo 1* e duas no *Grupo 2*. Os dados modificados, designados por Hemofilia modificados, encontram-se no Apêndice B. Iniciámos a análise representando sobre os dados os contornos de densidade constante, com o quadrado da distância de Mahalanobis igual a 5.99, que correspondem às elipses com 95% de probabilidade (aproximadamente). Além disso, na Figura 6.2 também foram sobrepostos os eixos das direcções principais. Como se observa, as cinco observações modificadas nos dois grupos, identificadas a vermelho, são as únicas que se encontram fora dos limites das elipses. Por outro lado, os eixos das *componentes principais* dos dois grupos não parecem verificar o modelo das *componentes principais comuns*. Esta constatação é também corroborada pelos valores das análises em *componentes principais* para cada grupo em separado, ver Tabela 6.2.



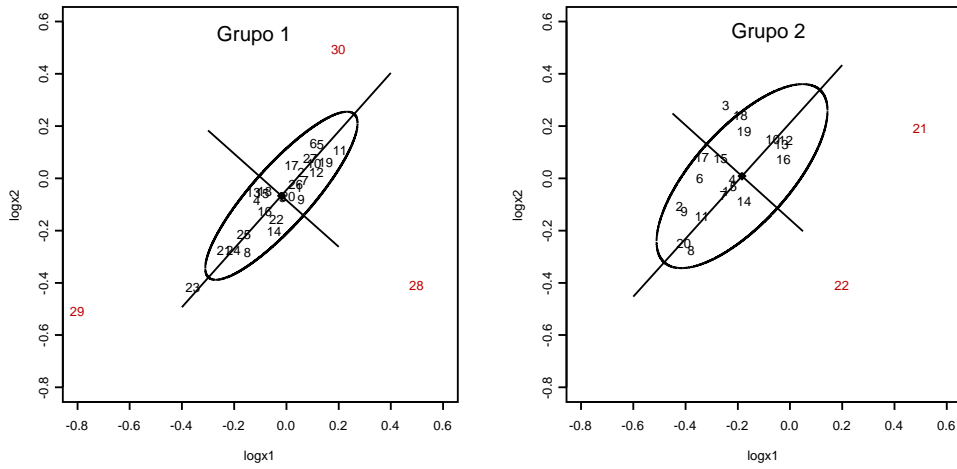
**Figura 6.2:** Elipses correspondentes a distância de Mahalanobis clássica constante e eixos das CP.

No entanto, utilizando a versão robusta da distância de Mahalanobis (Rousseeuw e van Zomeren, 1990), com as localizações e as dispersões estimadas com o RMCD, o panorama anterior é alterado, parecendo agora que ambos os grupos partilham as mesmas direcções principais, como se vê na Tabela 6.3 e na Figura 6.3.

Para confirmar a análise exploratória, a adequação do modelo CPC às matrizes de dispersão RMCD foi avaliada pelo teste de ajustamento de  $H_{CPC}$ . O valor observado

	<i>Grupo 1</i>		<i>Grupo 2</i>	
	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>1</sub>	CP <sub>2</sub>
valores próprios	0.070	0.018	0.052	0.027
vectors próprios				
$\log_{10}x_1$	-0.766	0.642	-0.904	0.427
$\log_{10}x_2$	-0.643	-0.766	-0.427	-0.904

**Tabela 6.2:** Análises em CP clássicas para cada grupo em separado.



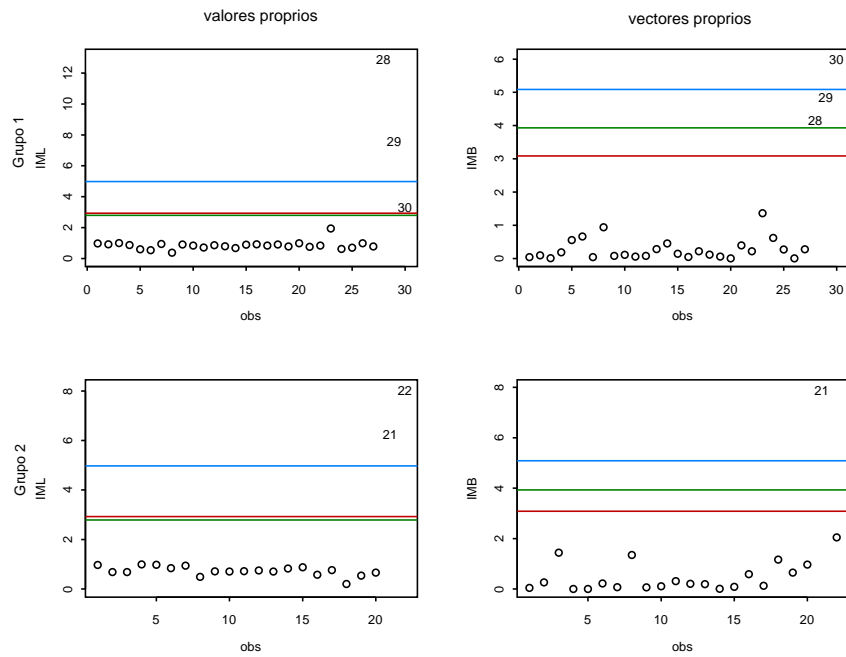
**Figura 6.3:** Elipses correspondentes a distância de Mahalanobis robusta constante e eixos das CP.

	<i>Grupo 1</i>		<i>Grupo 2</i>	
	CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>1</sub>	CP <sub>2</sub>
valores próprios	0.029	0.002	0.033	0.006
vectors próprios				
$\log_{10}x_1$	0.666	0.746	0.667	0.743
$\log_{10}x_2$	0.746	-0.666	0.743	-0.670

**Tabela 6.3:** Análises em CP robustas para cada grupo em separado.

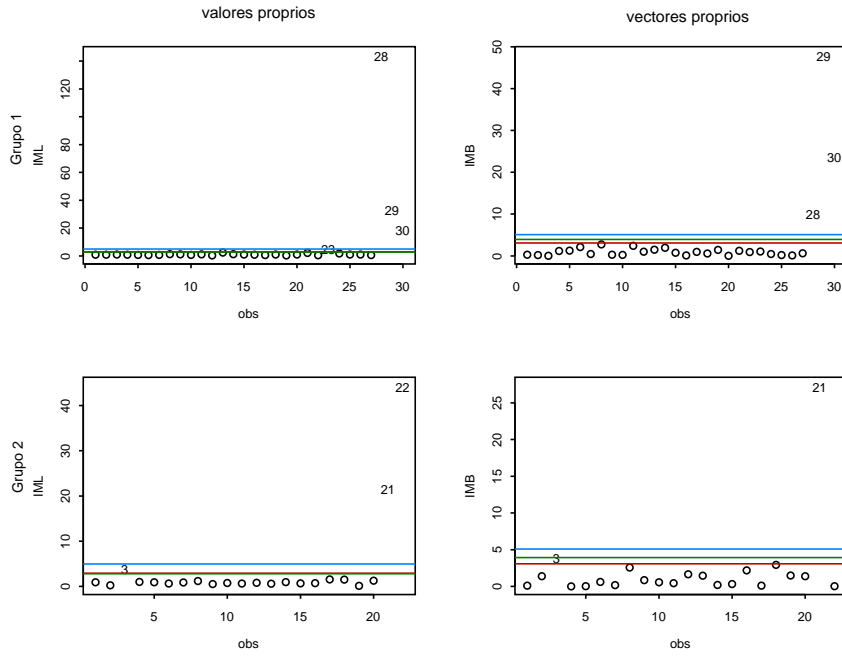
da estatística de teste, 0.025, conduziu a um *valor-p* da ordem dos 87%. Para as matrizes de covariâncias clássicas este valor é reduzido quase para metade, 47%. Parece assim ser razoável considerar os dados de hemofilia modificados para avaliar o comportamento das duas medidas de diagnóstico introduzidas na Secção 2. Assim, com base nos valores de  $IML_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$  e  $IMB_i(\mathbf{x}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ ,  $i = 1, 2$ , as Figuras 6.4 e 6.5

reportam a influência das observações na estimação dos valores e vectores próprios. Na Figura 6.4 a técnica utilizada na estimação de  $\hat{\beta}$  e  $\hat{\lambda}$  foi o “plug-in” das matrizes de covariâncias clássicas enquanto que na Figura 6.5 aplicaram-se matrizes robustas RMCD. Refira-se que a Figura 6.4 foi apresentada só para efeitos de comparação, pois a técnica utilizada não está de acordo com a que preconizamos. As Figuras 6.6 e 6.7 referem a estimação dos parâmetros de interesse com a metodologia usual de “projection-pursuit” e a escala estimada via MAD e via estimador-M. A estimação decorrente da metodologia de PP com a aplicação da função logaritmo encontra-se nas Figuras 6.8 e 6.9, com a escala estimada com o MAD e o estimador-M, respectivamente.

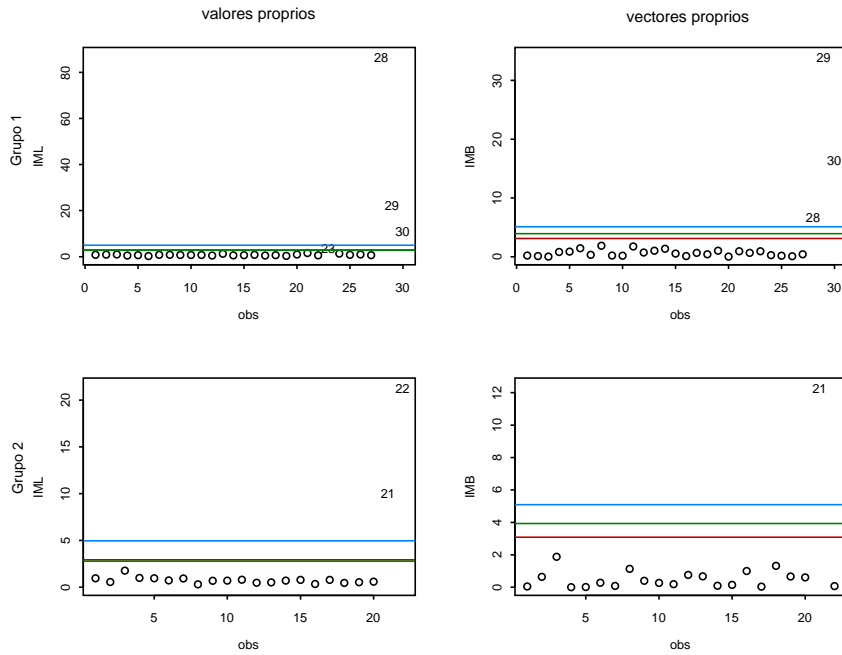


**Figura 6.4:** Detecção de observações discordantes/influentes com PI clássico. As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_{\lambda}$  e  $G_{\beta}$ .

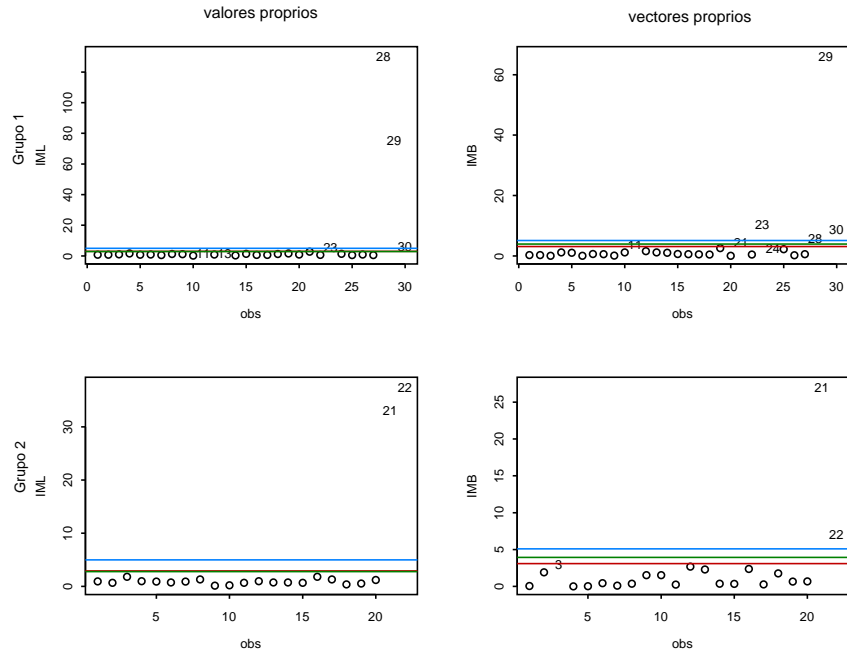
A análise das figuras apresentadas revela que as medidas IML e IMB detectam as mesmas observações, ou mais, que a distância de Mahalanobis robusta. No entanto, as duas medidas de diagnóstico introduzidas e a distância de Mahalanobis não estão a medir a mesma quantidade e podem, eventualmente, não ser tão concordantes na detecção. É evidente a vantagem da nossa abordagem, pois com ela é possível distinguir a influência das observações nos valores e vectores próprios, o que não acontece com a distância de Mahalanobis. Por exemplo, a observação 22 do *Grupo 2* é muito influente na estimação dos valores próprios desse grupo mas no entanto



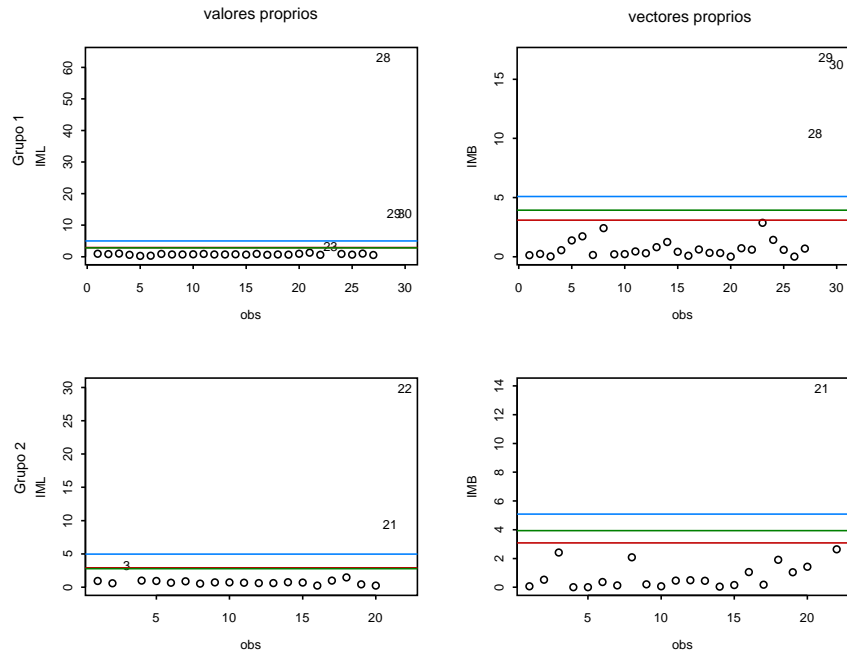
**Figura 6.5:** Detecção de observações discordantes/influentes com PI robusto. As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_\lambda$  e  $G_\beta$ .



**Figura 6.6:** Detecção de observações discordantes/influentes com  $PP_1$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_\lambda$  e  $G_\beta$ .

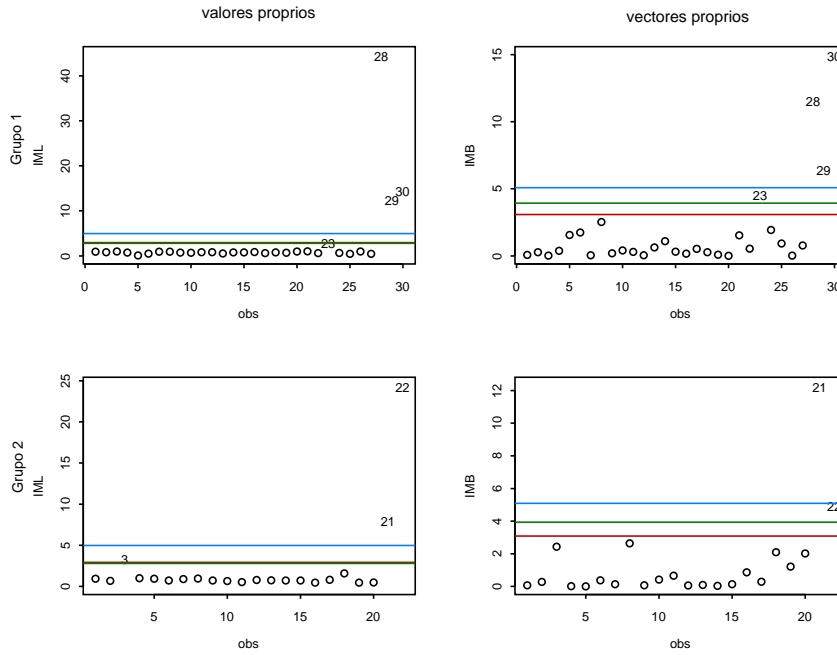


**Figura 6.7:** Detecção de observações discordantes/influentes com  $PP_2$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_\lambda$  e  $G_\beta$ .



**Figura 6.8:** Detecção de observações discordantes/influentes com  $LPP_1$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_\lambda$  e  $G_\beta$ .





**Figura 6.9:** Detecção de observações discordantes/influentes com  $LPP_2$ . As linhas a vermelho, verde e azul indicam, respectivamente, os percentis 95%, 97.5% e 99% de  $G_\lambda$  e  $G_\beta$ .

exerce pouca influência na estimação dos vectores próprios comuns. Por outro lado, nota-se que a alteração do processo de estimação dos parâmetros desconhecidos não parece, neste caso, influenciar os resultados da detecção (talvez porque os pontos introduzidos são claramente discordantes, em situações mais de “fronteira” já poderá não haver tanta concordância).

## Exemplo 2 (Variedades de Castanheiros)

Para estes dados o estudo foi iniciado com a construção de gráficos semelhantes aos das figuras anteriores. No entanto, pareceu-nos mais conveniente condensar a informação relevante em tabelas porque o elevado número de observações originou, por vezes, sobreposições e dificuldades na leitura. Assim, nas Tabelas 6.4, 6.5 e 6.6 registam-se as observações detectadas como possíveis “outliers” pelo IML e pelo IMB, com critério de registo baseado nos limites de confiança 95%, 97.5% e 99%. Na Tabela 6.4 são também indicadas as observações influentes para outros três métodos de diagnóstico, baseados na distância de Mahalanobis robusta com o RMCD, versões clássicas do IML e IMB e a distância de Mahalanobis clássica. Em todas as tabelas os números que identificam as observações foram ordenados por ordem crescente dos valores das correspondentes medidas, utilizando-se || para separar as observações que se encontram entre os percentis anteriormente referidos.

O “masking effect” é confirmado nos métodos clássicos, como se nota na Tabela 6.4. Apenas são detectadas 6 observações no grupo *Lada* e 2 no grupo *Longal*. Além disso, no grupo *Longal* não foi detectada a observação com maior valor da distância de Mahalanobis robusta. Em contrapartida, quando comparamos as observações consideradas influentes com as medidas IML e IMB robustas com o “plug-in” da matriz RMCD, com as indicadas pela distância de Mahalanobis robusta, observamos que os dois métodos concordam nas mais influentes. Notam-se algumas discrepân-

Método	<i>Lada</i>	<i>Longal</i>
IML-PI	29   94 79 77   24 12 87 88 30	29 33   14   34 42 39 44 32 22 1 6 24
IMB-PI	98 72 45 85 94 12   88   30	*   32   42 14 34 44 39 6 22 24 1
RMD	79 45 98   94   12 87 88 30	*   *   42 14 34 32 44 39 6 22 24 1
IML Clássico	24 79   12   87 88 30	32   *   *
IMB Clássico	*   *   30	24   *   *
MD	87   88   30	24   *   *
RMD: Distância de Mahalanobis Robusta,		MD: Distância de Mahalanobis.

**Tabela 6.4:** Observações detectadas como possíveis “outliers” pelo método PI e distância de Mahalanobis.

Método	<i>Lada</i>	<i>Longal</i>
IML-PP1,DD	45   30 87   88	6   *   44 22 1 24
IMB-PP1,DD	85 100 72 93 45   94 88 87   30	*   *   22 6 1 24
IML-PP2,DD	18 79 77 45 12   *   87 88 30	10 21   7   24 14 22 1
IMB-PP2,DD	99 93 88   85   30	45 39 42   24   14 22 1
IML-PP1,RD	93 79   45 4   87 30 88	14 24   *   22 1
IMB-PP1,RD	12 85 92   45   93 87 88 30	21 14   24   22 1
IML-PP2,RD	72 93   18 77 45   12 87 30 88	29 14 6 10 42 21   *   22 1 24
IMB-PP2,RD	70 88   99 85 72 93   45 30	*   42 21   14 22 1

**Tabela 6.5:** Observações detectadas como possíveis “outliers” pelo método PP.

Método	<i>Lada</i>	<i>Longal</i>
IML-LPP1,DD	98 87 70   94   88 30	*   14   45 22 1 6 24
IMB-LPP1,DD	72 30 93   45   88 87	14 10 21   42   6 45 22 1 24
IML-LPP2,DD	93 99 95 94 24   18 45   12 87 88 30	39 42 44   32 14   6 22 1 24
IMB-LPP2,DD	94 93 70 99 87   45 88 85 72   30	34 42   14 39   6 24 22 1
IML-LPP1,RD	79 72 99 45   77   87 88 30	*   21   24 22 1
IMB-LPP1,RD	70 12 82 99   72 93 85   87 45 88 30	*   14 10   24 21 22 1
IML-LPP2,RD	72 93   18 77 45   12 87 30 88	29 14 6 10 42 21   *   22 1 24
IMB-LPP2,RD	70 88   99 85 72 93   45 30	*   42 21   14 22 1

**Tabela 6.6:** Observações detectadas como possíveis “outliers” pelo método LPP.

cias para as observações perto dos limites de detecção mas isto é compreensível uma vez que (IML, IMB) e a distância de Mahalanobis robusta medem efeitos distintos.

Por outro lado, os resultados da detecção com IML e IMB construídos com a técnica de “projection–pursuit” apresentam algumas diferenças consoante o método e/ou o estimador de escala empregue. Eventualmente, a existência de “inliers” nas observações projectadas poderá contribuir para este facto. Contudo, a maior parte das observações com maior valor da distância de Mahalanobis robusta, e as detectadas com (IML, IMB) com “plug–in” robusto, foram também detectadas com a técnica de “projection–pursuit”.

Com os exemplos apresentados, julgamos que foi elucidado o comportamento das novas medidas para detecção de eventuais observações discordantes/influentes. Pensamos que a vantagem de apenas duas medidas foi notória, em especial no exemplo das variedades de castanheiros com dados num espaço de dimensão 5. Além disso, com (IML, IMB) é possível quantificar a influência que as observações exercem na estimação dos valores e dos vectores próprios. Este facto é uma mais valia em relação, por exemplo, à distância de Mahalanobis robusta.

# Capítulo 7

## Conclusões

Neste capítulo apresentamos um resumo com as conclusões mais relevantes da investigação desenvolvida. Além disso, também são sugeridos alguns tópicos para trabalho futuro. Algumas das novas questões indicadas para futura investigação surgem quer de variantes dos métodos utilizados nesta tese quer da aplicação da metodologia apresentada a outros problemas de análise multivariada. As outras, referem-se a problemas que consideramos importantes e que não foram tratados no trabalho aqui apresentado por limitações óbvias de espaço e tempo.

### 7.1 Resultados mais relevantes

Como referimos no Capítulo 1, o principal objectivo desta tese foi a construção de novos métodos robustos para a análise em *componentes principais comuns*. Neste sentido, e dado o papel fundamental da *função de influência* no contexto da estatística robusta, o nosso trabalho centrou-se, essencialmente, no estudo do comportamento infinitesimal dos funcionais relacionados com os parâmetros de interesse do modelo CPC. Além disso estas funções ainda permitem quantificar um outro conceito com papel preponderante na estatística, a eficiência assintótica dos estimadores. O estudo desenvolveu-se em duas abordagens robustas distintas, uma conduzida com “plug-in” (PI) e outra com “projection-pursuit” (PP).

Assim, foi no Capítulo 4 que apresentámos as *funções de influência parciais* para os funcionais relacionados com os estimadores obtidos por PI e PP. A análise destas funções permite concluir que:

- as *funções de influência parciais* dos funcionais valores e vectores próprios

obtidos por PI são limitadas quando se considera um funcional da matriz de dispersão com *função de influência* limitada;

- as *funções de influência parciais* dos funcionais valores próprios obtidos por PP são limitadas quando se considera um funcional univariado de escala com *função de influência* limitada;
- as *funções de influência parciais* dos funcionais vectores próprios comuns obtidos por PP são ilimitadas mesmo quando se considera um funcional univariado de escala com *função de influência* limitada.

Além disso, para o caso particular do modelo normal e para o funcional clássico  $\sigma^2(F) = \text{VAR}(F)$  resulta que:

- a variância assintótica dos estimadores obtidos por PI ou PP dos valores próprios iguala a variância assintótica dos estimadores de máxima verosimilhança;
- a variância assintótica dos estimadores PI dos vectores próprios comuns iguala a dos estimadores de máxima verosimilhança, enquanto que para os estimadores PP a variância assintótica iguala a dos estimadores construídos com a matriz de covariâncias combinada das  $k$  populações.

A conclusão que talvez mais influencie no processo de estimação é a falta de *B-robustez* dos estimadores dos vectores próprios comuns para o caso PP.

Face a este cenário, a investigação prosseguiu com o objectivo de tentar criar novos estimadores que mantivessem a qualidade *B-robusta* para os valores próprios e fossem mais resistentes, em algum sentido, na estimação dos vectores próprios comuns. Assim, no Capítulo 5 foram introduzidos novos estimadores que aplicam a técnica de PP e mais gerais que os considerados no Capítulo 4. Esta opção deveu-se, essencialmente, às vantagens dos métodos PP. Relembre-se que o método PP funciona mesmo com mais variáveis que observações e utiliza um processo sequencial na estimação dos vectores próprios comuns. Por outro lado, estes estimadores podem usufruir de boas propriedades de robustez global. O estudo incidiu sobre os estimadores resultantes da pré-aplicação de uma função positiva  $f$  geral, estritamente crescente e duas vezes diferenciável, à variabilidade das variáveis projectadas. Para os funcionais equivalentes aos estimadores PP gerais, derivámos as *funções de influência parciais* dos valores próprios e vectores próprios comuns. Destas funções conclui-se que:

- as *funções de influência parciais* dos funcionais valores próprios são limitadas quando se considera um funcional univariado de escala com *função de influência* limitada e não dependem de  $f$ ;
- as *funções de influência parciais* dos funcionais vectores próprios comuns são ilimitadas mesmo quando se considera um funcional univariado de escala com *função de influência* limitada e dependem de  $f$ .

Quanto às variâncias assintóticas destes estimadores, conclui-se para o caso particular do modelo normal e para o funcional clássico  $\sigma^2(F) = \text{VAR}(F)$  que:

- a variância assintótica dos estimadores PP gerais dos valores próprios iguala a variância assintótica dos estimadores de máxima verosimilhança;
- a variância assintótica dos estimadores PP gerais dos vectores próprios comuns depende da função  $f$ .

O caso particular da função  $f(t) = \ln(t)$  (estimadores LPP) foi estudado com maior detalhe por traduzir uma abordagem inspirada na de Flury (1984). As alterações, mais relevantes, foram verificadas nas variâncias assintóticas dos estimadores LPP dos vectores próprios. No caso particular do modelo proporcional, quando  $\Sigma_i = \Lambda_i = \rho_i \Lambda_1$ , estes estimadores atingem o valor mínimo da variância assintótica. Foi possível demonstrar que a função  $f(t) = \ln(t)$  é a única que minimiza a variância assintótica dentro de uma classe mais restrita de funções do tipo  $\mathcal{F} = \{f(t) : f'(t) = t^{\alpha-1}, \alpha \geq 0\}$ . Além disso, a eficiência assintótica dos estimadores LPP não depende da constante de proporcionalidade contrariamente ao que acontece com os estimadores PP usuais. O desempenho de vários estimadores LPP e PP foi comparado, gerando amostras normais e contaminadas de pequenas dimensões, através de um estudo de simulação de Monte Carlo. Foi também efectuada a estimação dos valores e vectores próprios comuns de dois conjuntos de dados reais. Desenvolveu-se também um novo método para a geração de direcções aleatórias (RD) de projecção das observações. As conclusões mais marcantes do estudo de simulação foram:

- nenhum dos estimadores utilizados se mostrou uniformemente melhor para a estimação dos valores próprios;
- os estimadores LPP mostraram-se vantajosos na estimação dos vectores próprios comuns, mesmo sem a validade do modelo proporcional;

- a nova alternativa de pesquisa das direcções de projecção não mostrou vantagens significativas face à proposta das direcções dos dados (DD) considerada por Boente e Orellana (2001), embora possa eventualmente ser uma boa alternativa para amostras mais pequenas ou quando o número de variáveis é elevado.

A detecção de observações discordantes/influentes é um passo importante em qualquer análise, mesmo quando são utilizados estimadores robustos. Neste sentido, no Capítulo 6 apresentámos uma proposta para a detecção de eventuais dados anómalos baseada na estimação dos eixos principais e seus comprimentos, i.e, na decomposição espectral das matrizes de dispersão. Esta alternativa é diferente das usuais medidas de diagnóstico, baseadas na distância de Mahalanobis, pois não utiliza directamente as matrizes de dispersão. Podemos salientar como vantagens das duas novas medidas de diagnóstico (IML, IMB) que:

- o número de gráficos a analisar é reduzido para  $2k$ , enquanto que quando se analisa em separado a influência exercida em cada parâmetro se constroem  $\left\{ p + \frac{1}{2} p(p-1) \right\} k$ ;
- é possível avaliar a influência das observações nos valores e vectores próprios separadamente;
- evitam o “masking effect”;
- detectaram tantas ou mais observações possivelmente discordantes/influentes que outras metodologias nos dois exemplos reais considerados.

Com base nos resultados apresentados nos Capítulos 5 e 6 e noutros estudos prévios, recomendamos que as medidas de diagnóstico sejam baseadas nos estimadores LPP com o estimador-M de escala. O estimador-M tem a grande vantagem de poder combinar razoavelmente a eficiência com o *alto ponto de ruptura*. Por outro lado, como já tinha sido referido por Boente e Orellana (2001), os estimadores PI são bastante sensíveis à contaminação quando esta é exercida na direcção do menor valor próprio.

## 7.2 Trabalho futuro

Para futura investigação, começamos por sugerir a avaliação de algumas modificações nos estimadores de “projection-pursuit” apresentados nesta tese. A fraca efi-

ciência das estimativas dos vectores próprios poderá estar eventualmente relacionada com a baixa eficiência dos estimadores de escala utilizados, MAD e estimador-M. Provavelmente, melhores resultados poderiam ser obtidos tomando-se um outra função  $\chi$  no estimador-M (Croux, 1994) ou utilizando um estimador-M ponderado só com 1 ou 2 passos (Rousseeuw e Croux, 1994). Outra alternativa também interessante de investigar é a resultante da aplicação do estimador-Q (Rousseeuw e Croux, 1992, 1993). Refira-se que no contexto da análise em *componentes principais*, Croux e Ruiz-Gazen (2000) concluíram que o estimador-Q originava bons resultados para as estimativas dos valores e vectores próprios com a técnica PP. Neste caso, a eficiência destes estimadores dos vectores próprios é igual a 67%, muito melhor que a alcançada com outros estimadores de escala robustos.

Contrariamente às nossa expectativas, a nova técnica RD de pesquisa de direcções não mostrou grandes vantagens no estudo desenvolvido. No entanto, pensamos que outra alternativa interessante pode ser criada com um processo híbrido do DD e RD, isto é, utilizando simultaneamente direcções aleatórias e as dos dados.

No estudo de simulação desenvolvido observou-se que o estimador PP baseado na função  $f(t) = \ln(t)$  em algumas situações, não conseguia por exemplo distinguir  $\beta_1$  de  $\beta_2$ . Pensamos que nestas circunstâncias o desempenho destes estimadores devia ser avaliado calculando a distância entre o espaço linear gerado por  $\beta_1$  e  $\beta_2$  e o gerado por  $\hat{\beta}_1$  e  $\hat{\beta}_2$ .

O algoritmo empregue na construção dos estimadores PP foi o utilizado por Boente e Orellana (2001) e que generaliza o algoritmo (CR) de Croux e Ruiz-Gazen (1996) para  $k$  populações. No entanto, mais recentemente Hubert *et al.* (2002) propõem um outro algoritmo designado por RAPCA e que os autores consideram mais rápido que o CR e mais estável em altas dimensões. Seria interessante desenvolver a generalização do algoritmo RAPCA para  $k$  populações. O RAPCA é composto por dois passos. No primeiro passo, o espaço dos dados é reduzido a um subespaço afim. Isto pode ser alcançado com a construção das *componentes principais* clássicas. No segundo passo, passo-R, é efectuada uma reflexão dos “scores”. Para isso, calcula-se o primeiro vector próprio  $\mathbf{b}_1$  do mesmo modo que o do algoritmo CR. Os “scores”  $\mathbf{t}_i^{(1)}$  são transformados por uma reflexão  $U_1 \left( \mathbf{t}_i^{(1)} \right) = \mathbf{t}_i^{(2)}$ , reflexão essa que quando aplicada a  $\mathbf{b}_1$  encontra o primeiro vector da base canónica de  $\mathbb{R}^p$ . De seguida projecta-se  $\mathbf{t}_i^{(1)}$  no complemento ortogonal de  $U_1(\mathbf{b}_1)$ . Isto é conseguido com a omissão da primeira coordenada de  $\mathbf{t}_i^{(2)}$ . Desta forma o segundo vector próprio  $\mathbf{b}_2^*$  é procurado em  $\mathbb{R}^{p-1}$  de acordo com o algoritmo CR, o qual por reflexão inversa é transportado para  $\mathbb{R}^p$ . O processo é repetido até se encontrarem todos os vectores



próprios.

A construção das *funções de influência parciais* para as medidas de diagnóstico seria também uma questão muito interessante.

Outro assunto relevante, do nosso ponto de vista, seria a construção de estimadores ponderados das *componentes principais comuns* em analogia com o que se faz para as *componentes principais* e para as matrizes de dispersão. Em relação a estas duas situações, estudos de vários autores revelam vantagens dos estimadores ponderados. A nossa sugestão é atribuir pesos às observações de acordo com a influência que elas exercem na estimação dos valores e vectores próprios. Assim, dada uma observação  $\mathbf{x}_{i\ell}$  sugerimos pesá-la com  $w_{i\ell} = w \left( IMB_i \left( x_{i\ell}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}} \right), IML_i \left( x_{i\ell}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}} \right) \right)$ , onde  $w : \mathbb{R}^2 \rightarrow \mathbb{R}^+$  é uma função “suave”, e posteriormente aplicar a técnica PI.

A derivação da distribuição assintótica dos estimadores de “projection–pursuit” das CPC é uma questão que se mantém em aberto e que merece ser estudada. Opinião idêntica temos relativamente à construção das *curvas de enviesamento máximo parciais* dos funcionais equivalentes.

Por último refira-se que as propostas robustas desenvolvidas nesta tese podem ser estendidas a outros tipos de aplicações. No trabalho de Orellana (1999) são referidos alguns temas onde é possível desenvolver uma abordagem equivalente à desta tese. Destes temas, salientamos o problema da *análise em variáveis canónicas comuns* estudado por Neuenschwander e Flury (1995) e por Gorla e Flury (1996). Em particular, a técnica utilizada para a construção das *funções de influência parciais* pode ser aplicada neste problema que envolve  $k$  populações (generalizando o trabalho de Croux e Dehon (2002) para  $k$  populações).

Esperamos que o trabalho desenvolvido nesta tese tenha contribuído, de alguma forma, para a divulgação da análise robusta em *componentes principais comuns*. Desejamos ainda que haja continuidade de investigação nesta área da estatística.

# Apêndice A Resultados de Álgebra

**Resultado 1:** Uma matriz  $\mathbf{A}$  de ordem  $p$  cujo espectro contém  $r$  valores próprios distintos com multiplicidades  $(\nu_1, \dots, \nu_r)$  é diagonalizável se e só se  $\sum_{j=1}^r \nu_j = p$ , o que é equivalente a dizer que a multiplicidade dos  $r$  ou  $(r - 1)$  valores próprios distintos de zero de  $\mathbf{A}$  é igual à característica de  $\mathbf{A}$ .

**Resultado 2:** Seja  $\mathbf{A}$  uma matriz  $(m \times p)$  e  $\mathbf{X}$  uma matriz  $(p \times k)$ . Se  $\mathbf{AX} = \mathbf{0}$  então, o espaço coluna de  $\mathbf{X}$  está contido no espaço nulo de  $\mathbf{A}$ , isto é,  $\mathcal{C}(\mathbf{X}) \subset \mathcal{N}(\mathbf{A})$ .

**Operador vec:** Seja  $\mathbf{V}$  uma matriz  $(r \times s)$ , então  $\text{vec}(\mathbf{V})$  é um vector coluna com  $(r \times s)$  linhas.

Assim, sendo

$$\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_s),$$

aplicando  $\text{vec}$  fica

$$\text{vec}(\mathbf{V}) = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_s \end{bmatrix}.$$

**Operador vech:** Seja  $\mathbf{V}$  uma matriz  $(p \times p)$ , simétrica ou não, então  $\text{vech}(\mathbf{V})$  é um vector coluna com  $p(p + 1)/2$  linhas, construído com os elemento que se encontram na triangular superior (ou inferior) desta matriz, ou seja

$$\text{vech}(\mathbf{V}) = (v_{11}, v_{12}, \dots, v_{1p}, v_{22}, \dots, v_{2p}, \dots, v_{(p-1)(p-1)}, v_{(p-1)p}, v_{pp})^T.$$



# Apêndice B Conjuntos de dados

## B.1 Dados Iris (dois grupos e duas variáveis)

<i>Versicolor</i>		<i>Virginica</i>	
<i>sepal length</i>	<i>sepal width</i>	<i>sepal length</i>	<i>sepal width</i>
7.0	3.2	6.3	3.3
6.4	3.2	5.8	2.7
6.9	3.1	7.1	3.0
5.5	2.3	6.3	2.9
6.5	2.8	6.5	3.0
5.7	2.8	7.6	3.0
6.3	3.3	4.9	2.5
4.9	2.4	7.3	2.9
6.6	2.9	6.7	2.5
5.2	2.7	7.2	3.6
5.0	2.0	6.5	3.2
5.9	3.0	6.4	2.7
6.0	2.2	6.8	3.0
6.1	2.9	5.7	2.5
5.6	2.9	5.8	2.8
6.7	3.1	6.4	3.2
5.6	3.0	6.5	3.0
5.8	2.7	7.7	3.8
6.2	2.2	7.7	2.6
5.6	2.5	6.0	2.2
5.9	3.2	6.9	3.2
6.1	2.8	5.6	2.8

(continua)

<i>sepal length</i>	<i>sepal width</i>
6.3	2.5
6.1	2.8
6.4	2.9
6.6	3.0
6.8	2.8
6.7	3.0
6.0	2.9
5.7	2.6
5.5	2.4
5.5	2.4
5.8	2.7
6.0	2.7
5.4	3.0
6.0	3.4
6.7	3.1
6.3	2.3
5.6	3.0
5.5	2.5
5.5	2.6
6.1	3.0
5.8	2.6
5.0	2.3
5.6	2.7
5.7	3.0
5.7	2.9
6.2	2.9
5.1	2.5
5.7	2.8

<i>sepal length</i>	<i>sepal width</i>
7.7	2.8
6.3	2.7
6.7	3.3
7.2	3.2
6.2	2.8
6.1	3.0
6.4	2.8
7.2	3.0
7.4	2.8
7.9	3.8
6.4	2.8
6.3	2.8
6.1	2.6
7.7	3.0
6.3	3.4
6.4	3.1
6.0	3.0
6.9	3.1
6.7	3.1
6.9	3.1
5.8	2.7
6.8	3.2
6.7	3.3
6.7	3.0
6.3	2.5
6.5	3.0
6.2	3.4
5.9	3.0

## B.2 Dados Iris Modificados (dois grupos e duas variáveis)

<i>Versicolor</i>		<i>Virginica</i>	
<i>sepal length</i>	<i>sepal width</i>	<i>sepal length</i>	<i>sepal width</i>
7.0	3.2	6.3	3.3
6.4	3.2	5.8	2.7
6.9	3.1	7.1	3.0
5.5	2.3	6.3	2.9
6.5	2.8	6.5	3.0
5.7	2.8	7.6	3.0
6.3	3.3	4.9	2.5
4.9	2.4	7.3	2.9
6.6	2.9	6.7	2.5
5.2	2.7	7.2	3.6
5.0	2.0	6.5	3.2
5.9	3.0	6.4	2.7
6.0	2.2	6.8	3.0
6.1	2.9	5.7	2.5
5.6	2.9	5.8	2.8
6.7	3.1	6.4	3.2
5.6	3.0	6.5	3.0
5.8	2.7	7.7	3.8
6.2	2.2	7.7	2.6
5.6	2.5	6.0	2.2
5.9	3.2	6.9	3.2
6.1	2.8	5.6	2.8
6.3	2.5	7.7	2.8
6.1	2.8	6.3	2.7
6.4	2.9	6.7	3.3
6.6	3.0	7.2	3.2
6.8	2.8	6.2	2.8
6.7	3.0	6.1	3.0
6.0	2.9	6.4	2.8
5.7	2.6	7.2	3.0
5.5	2.4	7.4	2.8
5.5	2.4	7.9	3.8
5.8	2.7	6.4	2.8
6.0	2.7	6.3	2.8
5.4	3.0	6.1	2.6
6.0	3.4	7.7	3.0
(continua)			

<i>sepal length</i>	<i>sepal width</i>
6.7	3.1
6.3	2.3
5.6	3.0
5.5	2.5
5.5	2.6
6.1	3.0
5.8	2.6
5.0	2.3
5.6	2.7
5.7	3.0
5.7	2.9
6.2	2.9
5.1	2.5
5.7	2.8

<i>sepal length</i>	<i>sepal width</i>
6.3	3.4
6.4	3.1
6.0	3.0
6.9	3.1
6.7	3.1
6.9	3.1
5.8	2.7
6.8	3.2
6.7	3.3
6.7	3.0
6.3	2.5
6.5	3.0
6.2	3.4
5.0	5.0

## B.3 Dados Hemofilia

*Grupo 1*

$\log_{10} x_1$	$\log_{10} x_2$
0.04922	-0.02687
0.05691	0.03342
-0.01323	-0.06550
-0.11351	-0.07572
0.13033	0.13988
0.10380	0.14302
0.07188	0.00000
-0.14874	-0.27572
0.05691	-0.07058
0.10721	0.06446
0.20683	0.11394
0.11727	0.03342
-0.12494	-0.04576
-0.04576	-0.19382
-0.09152	-0.05061
-0.08092	-0.11919
0.02119	0.05691
-0.08092	-0.04096
0.15229	0.06819
0.00860	-0.06048
-0.23657	-0.26761
-0.03621	-0.14874
-0.35655	-0.40894
-0.20066	-0.26761
-0.16115	-0.20761
0.03743	-0.01323
0.09342	0.08636
0.23300	0.19590
0.03743	0.12385
0.01284	0.11059

*Grupo 2*

$\log_{10} x_1$	$\log_{10} x_2$
-0.24466	-0.04067
-0.42318	-0.09981
-0.24529	0.28764
-0.22047	0.00455
-0.21539	-0.02191
-0.34470	0.00969
-0.25404	-0.05729
-0.37780	-0.26816
-0.40465	-0.11618
-0.06391	0.15694
-0.33510	-0.13676
-0.01493	0.15392
-0.03124	0.14001
-0.17402	-0.07764
-0.26421	0.08669
-0.02344	0.08038
-0.33525	0.08753
-0.18782	0.25096
-0.17443	0.18924
-0.40546	-0.24184
-0.24443	0.16137
-0.47837	0.02822



## B.4 Datos Hemofilia Modificados

<i>Grupo 1</i>		<i>Grupo 2</i>	
$\log_{10} x_1$	$\log_{10} x_2$	$\log_{10} x_1$	$\log_{10} x_2$
0.04922	-0.02687	-0.24466	-0.04067
0.05691	0.03342	-0.42318	-0.09981
-0.01323	-0.06550	-0.24529	0.28764
-0.11351	-0.07572	-0.22047	0.00455
0.13033	0.13988	-0.21539	-0.02191
0.10380	0.14302	-0.34470	0.00969
0.07188	0.00000	-0.25404	-0.05729
-0.14874	-0.27572	-0.37780	-0.26816
0.05691	-0.07058	-0.40465	-0.11618
0.10721	0.06446	-0.06391	0.15694
0.20683	0.11394	-0.33510	-0.13676
0.11727	0.03342	-0.01493	0.15392
-0.12494	-0.04576	-0.03124	0.14001
-0.04576	-0.19382	-0.17402	-0.07764
-0.09152	-0.05061	-0.26421	0.08669
-0.08092	-0.11919	-0.02344	0.08038
0.02119	0.05691	-0.33525	0.08753
-0.08092	-0.04096	-0.18782	0.25096
0.15229	0.06819	-0.17443	0.18924
0.00860	-0.06048	-0.40546	-0.24184
-0.23657	-0.26761	0.50000	0.20000
-0.03621	-0.14874	0.20000	-0.40000
-0.35655	-0.40894		
-0.20066	-0.26761		
-0.16115	-0.20761		
0.03743	-0.01323		
0.09342	0.08636		
0.50000	-0.40000		
-0.80000	-0.50000		
0.20000	0.50000		

## B.5 Dados Variedades de Castanheiros

*Lada*

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
17	14	10	19	17
10	16	15	20	21
21	11	11	19	19
16	13	16	16	19
20	11	12	15	18
23	12	11	17	16
18	13	15	19	17
17	11	11	17	17
18	9	11	16	16
16	10	12	16	16
22	10	12	15	16
24	16	14	19	20
17	13	12	18	17
15	12	13	16	17
13	11	11	18	20
16	13	14	18	18
15	10	12	16	18
18	17	15	21	20
14	15	12	17	16
10	15	15	20	19
17	10	10	15	15
13	15	13	19	18
20	13	9	17	16
16	9	12	17	15
21	7	9	13	16
13	13	12	17	16
12	16	13	21	20
23	8	7	14	15
17	7	10	18	17
11	15	14	20	12
10	14	14	21	18
14	10	13	16	17
14	13	13	19	18
12	12	13	20	19
10	13	14	19	21
16	12	11	18	17
18	8	11	16	15
10	10	11	17	18

*Longal*

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
36	22	25	24	28
38	12	12	15	16
35	14	11	17	16
35	15	16	17	18
22	22	19	23	20
15	14	15	14	15
22	17	21	18	22
27	13	15	17	18
20	19	21	20	23
26	8	9	13	14
20	21	18	23	20
26	20	22	22	22
26	15	16	16	17
32	19	23	21	25
23	19	19	20	21
30	18	15	20	18
35	12	14	18	18
24	20	19	22	20
37	10	11	16	16
24	19	19	20	21
29	7	8	12	14
32	22	25	24	28
27	20	19	20	21
39	24	21	27	25
17	21	21	22	21
19	16	18	18	21
25	18	19	19	20
35	15	15	16	19
22	21	22	23	20
26	18	17	20	17
23	18	16	20	19
18	15	16	21	21
40	19	19	20	21
38	20	19	20	23
40	14	16	18	18
20	20	18	22	21
33	11	11	14	15
30	10	12	15	16

(continua)

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
18	9	8	15	15
15	14	14	20	19
21	7	7	13	14
18	7	7	13	15
14	12	11	16	16
11	14	14	18	19
16	5	5	12	15
16	12	12	16	18
17	7	9	14	15
13	11	15	16	19
12	14	14	18	18
10	14	13	19	17
10	15	14	20	19
22	9	9	17	17
15	11	12	19	18
12	13	11	19	17
12	13	13	19	20
12	12	12	19	19
17	10	14	17	19
14	15	12	18	18
20	9	9	16	18
21	8	9	16	16
12	10	13	16	19
11	11	10	17	16
22	10	13	15	17
11	15	12	19	19
16	9	10	15	16
18	11	10	19	17
14	13	14	18	20
12	10	10	17	14
13	9	9	17	14
22	14	14	17	16
12	14	11	17	16
19	4	6	14	17
17	8	9	13	14
11	13	11	20	18
12	12	13	17	19
18	16	13	20	19
14	10	12	20	21

(continua)

(continuação)

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
28	19	20	21	25
25	16	17	18	20
17	20	20	21	20
10	14	15	19	17
37	16	15	18	17
22	16	19	17	17
34	22	18	24	19
28	16	16	19	18
17	18	20	22	22

(continuação)

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
14	11	10	19	18
18	10	8	13	16
15	12	11	18	18
13	16	14	20	19
19	5	5	15	15
18	8	7	15	13
19	9	9	16	18
21	4	4	15	14
17	16	16	19	20
24	16	16	20	21
23	18	17	20	21
23	8	8	16	16
15	11	9	16	14
10	15	12	19	19
18	6	6	12	13
18	6	5	12	12
23	10	10	20	19
22	6	7	17	17
19	12	12	18	17
21	7	8	15	17
19	11	10	21	18
20	15	16	17	18
25	7	6	16	14



# Apêndice C   Programas

## C.1 Programa LPP<sub>1</sub>

---

---

### PARÂMETROS DE ENTRADA

dat = “conjunto de dados com as variáveis por coluna.

A última coluna é a identificação do grupo”;

k = “número de grupos”;

ntent = “número de direcção aleatórias utilizadas na estimação dos vectores próprios comuns.”

---

---

### PROGRAMA PRINCIPAL

```
“ ppmaxmad” <- function(dat,k,ntent)
```

```
{
```

```
dat<-data.frame(dat)
```

```
n<-dim(dat)[1]
```

```
pt<-dim(dat)[2]
```

```
data <-(dat[, -pt])
```

```
p<-pt-1
```

```
library(Matrix)
```

```
soma<-NULL
```

```
buscamin<-NULL
```

```
indice<-NULL
```

```
agrupa<-dat[,pt]
```

```
if(n !=length(agrupa))
```

```
stop(“problema!!”)
```

---

```
(continua)
```

```

# avaliar a dimensão de cada grupo
g<-as.factor(agrupa)
counts<-tapply(rep(1,length(g)),g,sum)
# geração de  $\mathbf{Z}_i \sim N(\mathbf{0}_p, \mathbf{I}_p)$ 
varmad.grupo<-Matrix(NA,k,ntent)
nored<-rmvnorm(ntent, mean=rep(0,p), cov=diag(p),d=p)
nored<-as.Matrix(nored)
norma<-as.matrix(apply(nored,1,vecnorm))
normatrix<-matrix(rep(norma,p),ntent,p)
# possíveis  $\hat{\beta}_1$ 
beta1ini<-t(nored/normatrix)
for( j in 1:ntent)
{
# dados projectados nos possíveis  $\hat{\beta}_1$ 
xproj.grupo<-as.list(by(data,g,prod.fun,beta1ini[,j]))
# cálculo da escala MAD dos dados projectados
varmad.grupo[,j]<-as.matrix(sapply(xproj.grupo,mad.fun))
# função da escala combinada das variáveis projectas
parc<-log(varmad.grupo[,j])*counts
soma[j]<-colSums(parc)
}
# procura do máximo
buscamax<-rank(as.vector(soma))
for (i in 1:ntent)
{
if(buscamax[i]==n) indice<-i
}
lamb1<-varmad.grupo[indice]
lambaux<-lamb1
lamb<-Matrix(NA,k,p)
be1<-beta1ini[indice]
betaux<-be1

```

---

(continua)

```

be<-Matrix(NA,p,p)
be[,1]<-be1
lamb[,1]<-varmad.grupo[,indice]
# continuação do processo de construção da matriz Beta estimada
for (i in 2:p)
{
varmad.grupo<-Matrix(NA,k,ntent)
Aux<-(diag(p)-(betaux%*%t(betaux)))
noraux<-as.matrix(apply(Aux,2,vecnorm))
norauxmat<-matrix(rep(noraux,p),p,p)
norauxmat<-t(norauxmat)
Auxnorm<-Aux/norauxmat
norAuxnorm<-as.matrix(apply(Auxnorm,2,vecnorm))
betaini<-Auxnorm%*%betaini
norbeta<-as.matrix(apply(betaini,2,vecnorm))
norbetamatrix<-matrix(rep(norbeta,ntent),ntent,p)
norbetamatrix<-t(norbetamatrix)
betanorm<-betaini/norbetamatrix
for( j in 1:ntent)
{
xproj.grupo<-as.list(by(data,g,prod.fun,betanorm[,j]))
varmad.grupo[,j]<-as.matrix(sapply(xproj.grupo,mad.fun))
parc<-log(varmad.grupo[,j])*counts
soma[j]<-colSums(parc)
}
buscamax<-rank(soma)
for (l in 1:ntent)
{
if(buscamax[l]==n) indice<-l
}
lamb[,i]<-varmad.grupo[,indice]
lambaux<-cbind(lambaux,lamb[,i])
be[,i]<-betanorm[,indice]
betaux<-cbind(betaux,be[,i])
}

```

---

(continua)



```

betafinal<-betaux
lambfinal<-lambaux
print("matriz dos valores próprios estimados")
print((lambfinal)2)
print("matriz dos vectores próprios comuns estimados")
print(betafinal)
}

```

---

## FUNÇÕES AUXILIARES

```

“prod.fun”<-function(d1,d2)
{
prod<-d1%*%d2
}
# estimação da escala com o MAD
“mad.fun”<-function(d1)
{
out1<-mad(d1)
}

```

---

## C.2 Programa IML

---

---

### PARÂMETROS DE ENTRADA

dat = “matriz de dados ( $N \times (p + 1)$ ), com  $N = \sum_{i=1}^k n_i$  e  $p$  colunas com as observações das  $p$ -variáveis. A coluna  $p + 1$  é o vector com a identificação do grupo;”

beta = “matriz com as estimativas dos vectores próprios comuns ( $p \times p$ ), coluna  $j = j$ -ésimo vector próprio estimado sob a validade do modelo CPC;”

k = “número de grupos.”

l = “matriz com os valores próprios estimados para cada grupo,

linha  $i =$  valores próprios estimados do grupo  $i$ ;”

perc = “percentagem de observações a usar nas estimativas RMCD;”

b.95 = “estimativa do percentil de detecção a 95%;”

b.975 = “estimativa do percentil de detecção a 97.5%;”

b.99 = “estimativa do percentil de detecção a 99%.”

---

---

### PROGRAMA PRINCIPAL

```
“IMLhemoc” <- function(dat,beta,k,l,perc,b.95,b.975,b.99)
{
  dat<-data.frame(dat)
  beta<-as.matrix(beta)
  l<-as.matrix(l)
  n<-dim(dat)[1]
  pt<-dim(dat)[2]
  data <-(dat[, -pt])
  p<-pt-1
  agrupa<-dat[,pt]
  if(n !=length(agrupa))
  stop(“problema!”)
  # avaliar a dimensão de cada grupo
  g<-as.factor(agrupa)
  counts<-tapply(rep(1,length(g)),g,sum)
```

---

(continua)

```

# cálculo da média e matriz de dispersão RMCD
mcd.grupo<-as.list(by(data,g,mcdperc,perc=perc))
# centrar as observações com a média RMCD (y="observações cen-
tradas")
xcenter.grupo<-as.list(by(data,g,xcenter,perc=perc))
# calcular  $u = \beta^T \times y$ 
u<-lapply(xcenter.grupo,FUN=produto,beta)
# calcular  $u^2$ 
uquad<-lapply(u,FUN=quad.fun)
# cálculo de IML
for(i in 1:k)
{
IML<-apply(ifli.grupo,2,sum)
print("IML")
print(IML)
obs<-c(1:counts[i])
obs<-c(1:counts[i])
plot(obs,IML,xlim=c(1,counts[i]),ylim=c(0,max(b.99,sqrt(IML))),axes=T)
abline(b.95,0,col=8)
abline(b.975,0,col=4)
abline(b.99,0,col=6)
}
}

```

---



---

## FUNÇÕES AUXILIARES

```

# estimativa RMCD de localização e dispersão
"mcdperc"<-function(data,perc)
{
n<-nrow(data)
out1<-cov.mcd(data,quan=nperc)
}
# cálculo das observações estandardizadas
"xcenter"<-function(data,perc)

```

---

(continua)

```

{
n<-nrow(data)
out1<-cov.mcd(data,quan=percn)
mean.mcd<-out1$center
out2<-scale(data,center=out1$center,scale=F)
}
# esta função calcula  $u = \beta^T \times y$ 
“produto”<-function(xcenter.grupo,beta)
{
produ<-t(beta)%*%t(xcenter.grupo)
}
#função quadrado
“quad.fun”<-function(u)
{
uquad<-u^2
}

```

---

## C.3 Programa IMB

---

---

### PARÂMETROS DE ENTRADA

dat = “matriz de dados ( $N \times (p + 1)$ ), com  $N = \sum_{i=1}^k n_i$   
e  $p$  colunas com as observações das  $p$ -variáveis. A coluna  $p + 1$  é o vector identi-  
ficador do grupo;”  
beta = “matriz com as estimativas dos vectores próprios comuns ( $p \times p$ ),  
coluna  $j = j$ -ésimo vector próprio estimado sob a validade do modelo CPC;”  
k = “número de grupos.”  
l = “matriz com os valores próprios estimados para cada grupo,  
linha  $i$  = valores próprios estimados do grupo  $i$ ;”  
perc = “percentagem de observações a usar nas estimativas RMCD;”  
b.95 = “estimativa do percentil de detecção a 95%;”  
b.975 = “estimativa do percentil de detecção a 97.5%;”  
b.99 = “estimativa do percentil de detecção a 99%.”

---

---

### PROGRAMA PRINCIPAL

```
“IMBhemoc” <- function(dat,beta,k,l,perc,b.95,b.975,b.99)
{
  dat<-data.frame(dat)
  beta<-as.matrix(beta)
  l<-as.matrix(l)
  n<-dim(dat)[1]
  pt<-dim(dat)[2]
  data <-(dat[,-pt])
  p<-pt-1
  agrupa<-dat[,pt]
  if(n !=length(agrupa))
  stop(“problema!!”)
  # avaliar a dimensão de cada grupo
  g<-as.factor(agrupa)
  counts<-tapply(rep(1,length(g)),g,sum)
```

---

(continua)

```

# cálculo da média e matriz de dispersão RMCD
mcd.grupo<-as.list(by(data,g,mcdperc,perc=perc))
# centrar as observações com a média do RMCD (y="observações cen-
tradas")
xcenter.grupo<-as.list(by(data,g,xcenter,perc=perc))
# calcular  $u = \beta^T \times y$ 
u<-lapply(xcenter.grupo,FUN=produto,beta)
# calcular  $u^2$ 
uquad<-lapply(u,FUN=quad.fun)
#cálculo de IMB
for(i in 1:k)
{
func1<-(uquad[[i]]/rep(1[i,],counts[i]))
parc1<-(apply(func1,2,sum))^2
func2<-(func1)^2
parc2<-apply(func2,2,sum)
IMB<-parc1-parc2
print("IMB")
print(IMB)
obs<-c(1:counts[i])
plot(obs,IMB,xlim=c(1,counts[i]),ylim=c(0,max(b.99,sqrt(IMB))),axes=T)
abline(b.95,0,col=8)
abline(b.975,0,col=4)
abline(b.99,0,col=6)
}
}

```

---

## FUNÇÕES AUXILIARES

```

# estimativa RMCD de localização e dispersão
"mcdperc"<-function(data,perc)
{
n<-nrow(data)
out1<-cov.mcd(data,quan=n*perc)
}

```

---

(continua)

```

# cálculo das observações estandarizadas
"xcenter"<-function(data,perc)
{
n<-nrow(data)
out1<-cov.mcd(data,quan=perc*n)
mean.mcd<-out1$center
out2<-scale(data,center=out1$center,scale=F)
}
# esta função calcula  $u = \beta^T \times y$ 
"produto"<-function(xcenter.grupo,beta)
{
produ<-t(beta)%*%t(xcenter.grupo)
}
#função quadrado
"quad.fun"<-function(u)
{
uquad<-u2
}

```

---

# Referências Bibliográficas

- Achermann, R. R. e Cheverud, J. M. (2000). Phenotypic covariance structure in *Tomarins (genus saguinus)*: A comparison of variation patterns using matrix correlation and common principal components analysis. *American Journal of Physical Anthropology*, **111**, 489–501.
- Adrover, J. G. (1998). Minimax bias–robust estimation of the dispersion matrix of a multivariate distribution. *The Annals of Statistics*, **26**, 2301–2320.
- Airoidi, J. P. e Flury, B. N. (1988). An application of common principal component analysis to cranial morphometry of *Microtus californicus* and *M. ochrogaster* (Mammalia, Rodentia). *Journal Zool.*, **16**, 161–191.
- Airoidi, J. P. e Hoffmann, R. S. (1984). Age variation in voles (*microtus californicus*, *m. ochrogaster*) and its significance for systematic studies. Comunicação ocasional de *Museum of Natural History*, Lawrence, Kansas, University of Kansas, **111**, 1–45.
- Altman, E. I. (1968). Financial ratios, discriminante analysis and prediction of corporate bankruptcy. *Journal of Finance*, **23**, 589–609.
- Anderson, E. (1935). The irises of the Gaspé peninsula. *Bulletin of the American Iris Society*, **59**, 2–5.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, Nova Iorque, 1a. edição.
- Anderson, T. W. (1963). Asymptotic theory for principal components analysis. *Annals of Mathematical Statistics*, **34**, 122–148.
- Andrews, D. F., Brickel, P. J., Hampel, F. R. , Rogers, W. H. e Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, Nova Iorque.



- Arnold, S. J. e Phillips, P. C. (1999). Hierarchical comparison of generic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*. *Evolution*, **53**, 1516–1527.
- Barnett, V. e Lewis, T. (1978). *Outliers in Statistical Data*. Wiley, Chichester, 1a. edição.
- Barnett, V. e Lewis, T. (1984). *Outliers in Statistical Data*. Wiley, Chichester, 2a. edição.
- Barnett, V. e Lewis, T. (1994). *Outliers in Statistical Data*. Wiley, Chichester, 3a. edição.
- Berrendero, J. R. (1996). *Contribuciones a la Teoría de la Robustez Respecto al Sesgo*. Tese de Doutoramento (em espanhol), Universidad Carlos III de Madrid.
- Bessel, F. W. (1818). *Fundamenta Astronomiae*. Nicolovius, Königsberg.
- Billingsley, P. (1986). *Probability and Measure*. Wiley, Nova Iorque, 2a. edição.
- Boente, G. (1983). *Métodos Robustos para Componentes Principales*. Tese de Doutoramento (em espanhol), Universidad de Buenos Aires.
- Boente, G. (1987). Asymptotic theory for robust principal components. *Journal of Multivariate Analysis*, **21**, 67–78.
- Boente, G. e Orellana, L. (2001). A robust approach to common principal components. Em *Statistics in Genetics and in the Environmental Sciences*, L. T. Fernholz *et al.* (editores), Basel: Birkhauser, 117–147.
- Boente, G., Pires, A. M. e Rodrigues, I. M. (2002a). Influence functions and outlier detection under the common principal components model: A robust approach. *Biometrika*, **89**, 861–875.
- Boente, G., Pires, A. M. e Rodrigues, I. M. (2002b). General projection-pursuit estimators for common principal components model: Influence functions and Monte Carlo study. Trabalho submetido.
- Bouma, B. N., van der Klaauw, M. M., Veltkamp, J. J., Starkenburg, A. E., van Tilburg, N. H. e Hermans, J. (1975). Evaluation of the detection of hemophilia carriers. *Thrombosis Research*, **7**, 339–350.

- Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, **40**, 318–335.
- Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, **29**, 231–237.
- Campbell, N. A. (1981). Graphical comparison of covariance matrices. *Australian Journal of Statistics*, **23**, 21–37.
- Campbell, N. A. e Atchley, W. R. (1981). The geometry of canonical variate analysis. *Systematic Zoology*, **30**, 268–280.
- Chen, H. J., Gnanadesikan, R. e Kettenring, J. R. (1974). Statistical methods for grouping corporations. *Sankhyā, Series B*, **36**, 1–28.
- Critchley, F. (1985). Influence in principal components analysis. *Biometrika*, **72**, 627–636.
- Croux, C. (1994). Efficient high-breakdown M-estimators of scale. *Statistics and Probability Letters*, **19**, 371–379.
- Croux, C. e Dehon, C. (2002). Analyse canonique basee sur des estimateurs robustes de la matrice de covariance. *La Revue de Statistique Appliquee*, **2**, 5–26.
- Croux, C. e Haesbroeck, G. (1999a). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, **71**, 161–190.
- Croux, C. e Haesbroeck, G. (1999b). Empirical influence functions for robust principal component analysis. Em *1999 Proceedings of the Statistical Computing Section of the American Statistical Association*, 201–206.
- Croux, C. e Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*, **87**, 603–18.
- Croux, C. e Haesbroeck, G. (2002). Maxbias curves of location estimators based on subranges. *Journal of Nonparametric Statistics*, **14**, 295–306.
- Croux, C. e Ruiz-Gazen, A. (1996). A fast algorithm for robust principal components based on projection-pursuit. Em *Compstat: Proceedings in Computational Statistics*, A. Prat (editor), Heidelberg: Physica-Verlag, 211–217.

- Croux, C. e Ruiz-Gazen, A. (2000). High breakdown estimators for principal components: The projection-pursuit approach revisited. Trabalho submetido. Université Libre de Bruxelles.
- Cui, B. e H., He, X. e Kai, Ng, K. W. (2002). Asymptotic distribution of principal components based on robust dispersion. Trabalho submetido.
- Daudin, J. J., Duby, C. e Trecourt, P. (1988). Stability of principal components analysis studied by the bootstrap method. *Statistics*, **19**, 241–258.
- Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, **15**, 1269–1292.
- Davies, P. L. (1992). The asymptotics of Rousseeuw’s minimum volume ellipsoid. *The Annals of Statistics*, **20**, 1828–1843.
- Devlin, S. J., Gnanadesikan, R. e Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, **76**, 354–362.
- Donoho, D. L. (1982). *Breakdown Properties of Multivariate Location Estimators*. Tese de doutoramento (em inglês), Universidade de Harvard.
- Donoho, D. L. e Huber, P. J. (1983). The notion of breakdown point. Em *A Festschrift for Erich L. Lehmann*, P. J. Bickel *et al.* (editores), Wadsworth, Belmont, California, 157–187.
- Donoho, D. L., Johnstone, I., Rousseeuw, P. e Stahel, W. (1985). Comentários a “Projection-pursuit” de P. J. Huber, *Annals of Statistics*, **13**, 496–500.
- Efron, B. e Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54–77.
- Eriksen, P. S. (1987). Proportionality of covariance matrices. *The Annals of Statistics*, **15**, 732–748.
- Federer, W. T. (1951). Testing proportionality of covariance matrices. *Annals of Mathematical Statistics*, **22**, 102–106.
- Fernholz, L. T. (1983). *von Mises Calculus for Statistical Functionals*. Lecture Notes in Statistics 19, Springer-Verlag, Nova Iorque.

- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.
- Flury, B. N. (1983a). Some relations between the comparison of covariance matrices and principal component analysis. *Computational Statistics and Data Analysis*, **1**, 97–109.
- Flury, B. N. (1983b). A generalization of principal components analysis to  $k$  groups. *Relatório Técnico 83-14*, Departamento de Estatística, Universidade de Purdue.
- Flury, B. N. (1984). Common principal components in  $k$  groups. *Journal of the American Statistical Association*, **79**, 892–898.
- Flury, B. N. (1986a). Proportionality of  $k$  covariance matrices. *Statistics and Probability Letters*, **4**, 29–33.
- Flury, B. N. (1986b). Asymptotic theory for common principal components analysis. *The Annals of Statistics*, **14**, 418–430.
- Flury, B. N. (1987a). Two generalizations of the common principal components model. *Biometrika*, **74**, 59–69.
- Flury, B. N. (1987b). A hierarchy of relationships between covariance matrices. Em *Advances in Multivariate Analysis*, A. K. Gupta (editor), Reidel Publishing Company, Boston, 31–43.
- Flury, B. N. (1988). *Common Principal Components and Related Multivariate Models*. Wiley, Nova Iorque.
- Flury, B. N. e Constantine, G. (1985). Algorithm AS 211, the F-G diagonalization algorithm. *Applied Statistics*, **34**, 177–183.
- Flury, B. N. e Gautschi, W. (1984). An algorithm for simultaneous orthogonal transformation of several positive definite matrices to nearly diagonal form. *Relatório Técnico 84-1*, Departamento de Estatística, Universidade de Purdue.
- Flury, B. N. e Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing*, **7**, 177–183.

- Flury, B. N. e Riedwyl, H. (1983). *Angewandte Multivariate Statistik*. Gustav Fischer, Stuttgart e Nova Iorque.
- Flury, B. N. e Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach*. Chapman e Hall, Londres.
- Friedman, J. H. e Tukey, J. W. (1974). A projection–pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, **C-23**, 881–890.
- Gauss, C. F. (1880). Bestimmung der genauigkeit der beobachtungen. *Zeitschrift für Astronomie und verwandte Wissenschaften*, (Março e Abril 1816). Reproduzido em *Werke*, **8**, Göttingen: Dieterichsche Universitäts–Drucherei, 109–117.
- Gervini, D. (2002). The influence function of the Stahel–Donoho estimator of multivariate location and scatter. *Statistics and Probability Letters*, **60**, 425–435.
- Gnanadesikan, R. e Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, **28**, 81–124.
- Goria, M. N. e Flury, B. N. (1996). Common canonical variates in  $k$  independent groups. *Journal of the American Statistical Association*, **91**, 1735–1742.
- Guttman, I., Kim, D. Y. e Olkin, I. (1983). Statistical inference for constants of proportionality. Em *Multivariate Analysis - VI*, P. R. Krishnaiah (editor), North Holland, Nova Iorque, 257–280.
- Hampel, F. R. (1968). *Contributions to the Theory of Robust Estimation*. Tese de Doutorado (em inglês), University of California, Berkeley.
- Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, **42**, 1887–1896.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383–393.
- Hampel, F. R. (1991). Between robustness and diagnostics. Em *Directions in Robust Statistics and Diagnostics, Part I*, W. Stahel e S. Weisberg (editores), Springer–Verlag, Nova Iorque, 121–130.
- Hampel, F. R. (2000). Robust inference. *Relatório Técnico* **93**, Eidgenössische Technische Hochschule (E.T.H.), Zurique, Suíça.

- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. e Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, Nova Iorque.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, Nova Iorque.
- Healy, M. J. R. (1968). Multivariate normal plotting. *Applied Statistics*, **17**, 157–161.
- Hawkins, D. M. (1980). *Identification of Outliers*. Chapman e Hall, Londres.
- Hermans, J. e Habbema, J. D. F. (1975). Comparison of five methods to estimate posterior probabilities. *EDV in Medizin und Biologie*, **6**, 14–19.
- Hodges, J. L. Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. Em *Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, University of California Press, Berkeley, California, 163–186.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, **24**, 417–441, 498–520.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321–327.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.
- Huber, P. J. (1965). A robust version of the probability ratio test. *Annals of Mathematical Statistics*, **36**, 1753–1758.
- Huber, P. J. (1968). Robust confidence limits. *Z. Wahrsch. Verw. Geb.*, **10**, 269–278.
- Huber, P. J. (1972). Robust statistics: A review. *Annals of Mathematical Statistics*, **43**, 1041–1067.
- Huber, P. J. (1981). *Robust Statistics*. Wiley, Nova Iorque.
- Huber, P. J. (1985). Projection-pursuit. *The Annals of Statistics*, **13**, 435–475.
- Huber, P. J. e Strassen, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *The Annals of Statistics*, **1**, 251–263; **2**, 223–224.

- Hubert, M. e Rousseeuw, P.J. (2002). ROBPCA: a new approach to robust principal component analysis. Trabalho submetido.
- Hubert, M., Rousseeuw, P.J. e Verboven, S. (2002). A fast method for robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **60**, 101-111.
- Jaupi, L. e Saporta, G. (1993). Using the influence function in robust principal components analysis. Em *New Directions in Statistical Data Analysis and Robustness*, S. Morgenthaler *et al.* (editores). Birkhauser Verlag AG, Basel, 147-156.
- Jolicoeur, P. (1963). The degree of generality of robustness in *Martes Americana*. *Growth*, **27**, 1-27.
- Jolicoeur, P. e Mosimann, J. E. (1960). Size and shape variation in the painted turtle: a principal component analysis. *Growth*, **24**, 339-354.
- Jöreskog, K. G. e Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models*, Abt Associates Inc, Cambridge.
- Kallianpur, G. e Rao, C. R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhya A*, **15**, 331-342.
- Kariya, T. e Eaton, M. L. (1977). Robust tests for spherical symmetry. *The Annals of Statistics* **5**, 206-215.
- Kendall, M. G. e Buckland, W. R. (1971). *A Dictionary of Statistical Terms*. Oliver e Boyd, Edimburgo, 3a. Edição.
- Kent, J. T. e Taylor, D. E. (1991). Redescending  $M$ -estimates of multivariate location and scatter. *The Annals of Statistics*, **19**, 2102-2119.
- Keramidas, E. M., Devlin, S. J. e Gnanadesikan, R. (1987). A graphical procedure for comparing the principal components of several covariance matrices. *Communications in Statistics B-Computation and Simulation*, **16**, 161-191.
- Khatri, C. G. (1967). Some distribution problems connected with the characteristic roots of  $\mathbf{S}_i \mathbf{S}_i^{-1}$ . *Annals of Mathematical Statistics*, **38**, 944-948.
- Kim, D. Y. (1971). Statistical inference for constants of proportionality between covariance matrices. *Relatório Técnico* **59**, Stanford University.

- Klingenberg, C. P., Neuenschwander, B. E. e Flury, B. N. (1996). Ontogeny and individual variation: Analysis of patterned covariance matrices with common principal components. *Systematic Biol.*, **45**, 135–137.
- Klingenberg, C. P. e Zimmermann, M. (1992). Static, ontogenetic, and evolutionary allometry: A multivariate comparison in nine species of water striders. *American Naturalist*, **140**, 601–620.
- Krzanowski, W. J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, **74**, 703–707.
- Krzanowski, W. J. (1982). Between-groups comparison of principal components—some sampling results. *Journal of Statist. Comput. Simul.*, **15**, 141–154.
- Krzanowski, W. J. (1984). Principal component analysis in the presence of group structure. *Applied Statistics*, **33**, 164–168.
- Krzanowski, W. J. (1990). Between-groups analysis with heterogeneous covariance matrices: The common principal components model. *Journal of Classification*, **7**, 81–98.
- Kruskal, J. B. (1969). Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new “index of condensation”. Em *Statistical Computation*, R. C. Milton e J. A. Nelder (editores). Academic, Nova Iorque.
- Kruskal, J. B. (1972). Linear transformation of multivariate data to reveal clustering. Em *Multidimensional Scaling: Theory and Application in the Behavioral Sciences, I, Theory*. Seminar Press, Nova Iorque e Londres.
- Li, G. e Chen, Z. (1985). Projection–pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *Journal of the American Statistical Association*, **80**, 759–766.
- Lopuhaä, H. P. (1989). On the relation between S–estimators and M–estimators of multivariate location and covariance. *The Annals of Statistics*, **17**, 1662–1683.
- Lopuhaä, H. P. e Rousseeuw, P. J. (1991). Breakdown point properties of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, **19**, 229–248.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science, India*, **2**, 49–55.



- Manly, B. F. J. (1985). *The Statistics of Natural Selection on Animal Populations*. Chapman e Hall, Londres.
- Manly, B. F. J. e Rainer, J. C. W. (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika*, **74**, 841–847.
- Maronna, R. A. (1976). Robust  $M$ -estimators of multivariate location and scatter. *The Annals of Statistics*, **4**, 51–67.
- Maronna, R. A. e Yohai, V. J. (1995). The behaviour of the Stahel-Donoho robust multivariate estimator. *Journal of the American Statistical Association*, **90**, 330–341.
- Matusita, K. (1956). Decision rule, based on the distance, for the classification problem. *Annals of the Institute of Statistical Mathematics*, **8**, 67–77.
- Morrison, D. F. (1990). *Multivariate Statistical Methods*. McGraw-Hill, Nova Iorque.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, Nova Iorque.
- Neuenschwander, B. E. e Flury, B. N. (1995). Common canonical variates. *Biometrika*, **82**, 553–560.
- Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best results. *Am. J. Math.*, **8**, 343–366.
- Newcomb, S. (1895). Astronomical constants (the elements of the four inner planets and the fundamental constants of astronomy). Suplemento de *American Ephemeris and Nautical Almanac for 1897*. Estados Unidos, Washington, D. C..
- Novi Inverardi, P. L. e Flury, B. N. (1992). Robust estimation of common principal components. *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Social*, **14**, 49–79.
- Oliveira, I. (1995). *Variedades de Castanheiros em Trás-os-Montes. Uma Análise em Componentes Principais dos Caracteres Morfológicos da Folha*. Tese de mestrado, Universidade de Lisboa.
- Orellana, L. (1999). *Métodos Robustos para el Análisis de Componentes Principales Comunes*. Tese de Mestrado (em espanhol), Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires.

- Pack, P., Jolliffe, I. T. e Morgan, B. J. T. (1988). Influential observations in principal components analysis: a case study. *Journal of Applied Statistics*, **15**, 37–50.
- Patak, Z. (1991). *Robust Principal Components*. Tese de Doutorado (em inglês), University of British Columbia, Vancouver.
- Pearson, K. (1901). On lines and planes of closet fit to systems of point in space. *Phil. Mag.*, **2**, 559–572.
- Phillips, P. C. (1994). *CPC-Common Principal Components Analysis Program*. Universidade do Texas, Arlington.
- Phillips, P. C. e Arnold, S. J. (1999). Hierarchical comparison of generic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution*, **53**, 1506–1515.
- Pillai, K. C. S., Al-Ani, S. e Jouris, G. M. (1969). On the distribution of the ratio of the roots of a covariance matrix and Wilk's criterion for tests of three hypotheses. *Annals of Mathematical Statistics*, **40**, 2033–2040.
- Pires, A. M. (1990). *Estimação Robusta e sua Aplicação a Componentes Principais*. Tese de Mestrado, Instituto Superior Técnico, UTL, Lisboa.
- Pires, A. M. (1995). *Análise Discriminante: Novos Métodos Robustos de Estimação*. Tese de Doutorado, Instituto Superior Técnico, UTL, Lisboa.
- Pires, A. M. e Branco, J. (2002). Partial influence functions. *Journal of Multivariate Analysis*, **83**, 451–468.
- Pison, G., Rousseeuw, P. J., Filzmoser, P. e Croux, C. (2000). A robust version of principal factor analysis. Em *Compstat: Proceedings in Computational Statistics*, J. Bethlehem e P. van der Heijden (editores), Heidelberg: Physica-Verlag, 385–390.
- Portnoy, S. e He, X. (2000). A robust journey in the new millennium. *Journal of the American Statistical Association*, **95**, 1331–1335.
- Prohorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probab. Appl.*, **1**, 157–214.
- Quenouille, M. H. (1956). Notes in bias estimation. *Biometrika*, **43**, 353–360.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, Nova Iorque, 2a. edição.

- Rao, C. R. (1983). Likelihood ratio tests for relationships between covariance matrices. Em *Studies in Econometrics, Time Series and Multivariate Statistics*. S. Karlin *et al.* (editores). Academic Press, Nova Iorque, 529–543.
- Rousseeuw, P. J. (1981). *New Infinitesimal Methods in Robust Statistics*. Tese de Doutorado (em inglês), Free University of Brussels.
- Rousseeuw, P. J. (1983). Multivariate estimation with high breakdown point. Trabalho apresentado no Fourth Pannonian Symposium on Mathematical Statistics and Probability. “Abstract” em *IMS Bulletin*, 1983.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871–881.
- Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. Em *Mathematical Statistics and Applications*, W. Grossmann *et al.* (editores), Reidel Publishing Company, Dordrecht (coproduzido por Akadémiai Kiadó, Budapeste), 283–297.
- Rousseeuw, P. J. e Croux, C. (1992). Explicit scale estimators with high breakdown point. Em *L1-Statistical Analysis and Related Methods*. Y. Dodge (editor). Amesterdão, Holanda.
- Rousseeuw, P. J. e Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, **88**, 1273–1283.
- Rousseeuw, P. J., e Croux, C. (1994). The bias of  $k$ -step M-estimators. *Statistics and Probability Letters*, **20**, 411–420.
- Rousseeuw, P. J., e Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, Nova Iorque.
- Rousseeuw, P. J. e Ronchetti, E. (1981). Influence curves of general statistics. *Journal of Computational and Applied Mathematics*, **7**, 161–166.
- Rousseeuw, P. J. e van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**, 212–223.
- Rousseeuw, P. J. e van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, **85**, 633–639.
- Rousseeuw, P. J. e Yohai, V. J. (1984). Robust regression by means of S-estimators. Em *Robust and nonlinear time series analysis*, J. Franke *et al.* (editores) Lecture Notes in Statistics, **26**, 256–272, Springer Verlag, Nova Iorque.

- Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, Nova Iorque.
- Schott, J. R. (1988). Common principal components subspace in two groups. *Biometrika*, **75**, 229–236.
- Schott, J. R. (1991). Some tests for common principal components subspaces in several groups. *Biometrika*, **78**, 771–777.
- Schott, J. R. (1998). Estimating correlation matrices that have common eigenvectors. *Computational Statistics and Data Analysis*, **27**, 445–459.
- Sengupta, S. e Boyle, J. S. (1998). Using common principal components for comparing GCM simulations. *Journal of Climate*, **11**, 816–830.
- Shi, L. (1997). Local influence in principal components analysis. *Biometrika*, **84**, 175–86.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, Londres.
- Stahel, W. A. (1981). *Robust Estimation: Infinitesimal Optimality and Covariance Matrix Estimators*. Tese de doutoramento (em alemão), ETH, Zurique.
- Staudte, R. G. e Sheather S. J. (1990). *Robust Estimation and Testing*. Wiley, Nova Iorque.
- Steppan, S. J. (1997). Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal components analysis. *Evolution*, **51**, 517–586.
- Tanaka, Y. (1988). Sensitivity analysis in principal components analysis: influence on the subspace spanned by principal components. *Communications in Statistics A*, **17**, 3157–3175.
- Tukey, J. W. (1958). Bias and confidence in not–quite large samples (Abstract). *Annals of Mathematical Statistics*, **29**, 614.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Em *Contributions to Probability and Statistics*. I. Olkin *et al.* (editores). Stanford University Press, Stanford, Califórnia.
- Tukey, J. W. (1970–71). *Exploratory Data Analysis* (Edição preliminar). Addison–Wesley, Reading, Mass.

- Tyler, D. E. (1983). The asymptotic distribution of principal components roots under local alternative to multiple roots. *The Annals of Statistics*, **11**, 1232–1242.
- Tyler, D. E. (1986). Breakdown properties of the M-estimators of multivariate scatter. *Relatório Técnico*, Department of Statistics, Rutgers University.
- Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, **15**, 234–251.
- Tyler, D. E. (1994). Finite-sample breakdown points of projection-based multivariate location and scatter statistics. *The Annals of Statistics*, **22**, 1024–1044.
- von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, **18**, 309–348.
- Xie, Y. L., Wang, J. H., Liang, Y. Z., Sun, L. X., Song, X. H. e Yu, R. Q. (1993). Robust principal components analysis by projection-pursuit. *J. Chemom.*, **7**, 527–541.
- Yvan, K. H. e Bentler, P. M. (1994). Test of linear trend in eigenvalues of  $k$  covariance matrices with applications in common principal components analysis. *Commun. Statist.-Theory Meth.*, **23**, 3141–3156.
- Zimmermann, M. (1987). Die larven der schweizerischen *Gerris*-Arten (Hemipters, Gerridae). *Revue Suisse de Zoologie*, **94**, 593–624.
- Zimmerman, T. S., Ratnoff, O. D. e Littel, A. S. (1971). Detection of carriers of classic hemophilia using an imonulologic assay for antihemophilia factor (factor VIII). *The Journal of Clinical Investigation*, **50**, 255–258.
- Zuo, Y., Cui, H. e He, X. (2001). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. Trabalho submetido.