



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

JHONNATA BEZERRA DE CARVALHO

MÉTODOS DE CLASSIFICAÇÃO E BONDADDE DE AJUSTE NA ANÁLISE DE
FORMAS PLANAS

Recife

2019

JHONNATA BEZERRA DE CARVALHO

**MÉTODOS DE CLASSIFICAÇÃO E BONDADDE DE AJUSTE NA ANÁLISE DE
FORMAS PLANAS**

Tese apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutor em Estatística.

Área de Concentração: Estatística Aplicada

Orientador: Prof. Dr. Getúlio José Amorim do Amaral

Coorientador: Prof. Dr. Alex Dias Ramos

Recife

2019

Catálogo na fonte
Bibliotecária Mariana de Souza Alves CRB4-2106

C331m Carvalho, Jhonnata Bezerra de
Métodos de classificação e bondade de ajuste na análise de
formas planas – 2019.
229f.: il., fig., tab.

Orientador: Getúlio José Amorim do Amaral
Tese (Doutorado) – Universidade Federal de Pernambuco.
CCEN, Estatística. Recife, 2019.
Inclui referências e apêndices.

1. Estatística Aplicada. 2. Aprendizado supervisionado. 3.
Bingham complexa. 4. Teste de bondade de ajuste. I. Amaral,
Getúlio José Amorim do (orientador). II. Título.

310

CDD (22. ed.)

UFPE-MEI 2019-146

JHONNATA BEZERRA DE CARVALHO

**MÉTODOS DE CLASSIFICAÇÃO E BONDADE DE AJUSTE NA ANÁLISE DE
FORMAS PLANAS**

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

Aprovada em: 11 de outubro de 2019.

BANCA EXAMINADORA

Prof.(º) Getúlio José Amorim do Amaral
UFPE

Prof.(ª) Renata Maria Cardoso Rodrigues de Souza
UFPE

Prof.(º) Telmo de Menezes e Silva Filho
UFPB

Prof.(ª) Fernanda De Bastiani
UFPE

Prof.(º) Abraão David Costa do Nascimento
UFPE

À minha família:

Oswaldo, Filomena, Valdir, Solange, Jeferson,

John Lennon e Inara.

AGRADECIMENTOS

A Deus por ter me dado o dom da vida e iluminado o meu caminho até aqui.

Agradeço ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco. Agradeço ao meu orientador Getúlio José Amorim do Amaral e coorientador Alex Dias Ramos, por terem me guiado durante todo o tempo do doutorado, pela dedicação, companheirismo e conselhos.

Agradeço aos meus pais, Valdir e Solange, pelo apoio e por terem acreditado em mim. Aos meus irmãos Jeferson e John Lennon.

Agradeço aos meus avós, Osvaldo e Filomena, por terem me criado, educado e me direcionado. Agradeço a todos os meus familiares e amigos que me ajudaram direta e indiretamente, em especial, Emerson, Jean, Diego, Ezequiel, Géssica, Laís, Moizés, Felipe e Rumenick.

Gostaria de agradecer imensamente à minha companheira Inara, pois sem a sua ajuda eu não teria conseguido. Obrigado por seu amor, carinho, paciência e dedicação. Agradeço também à sua família, que me acolheu com muito carinho, em especial: Inelda, Valtomiro, Francinaldo, Socorro, Victor, Neto e Maria Isis.

Agradeço aos meus afilhados, Victor e Vicente, pela alegria que me proporcionam.

Agradeço ao professor André Pinho, meu orientador na graduação e mestrado, no qual eu me espelho para ser um bom professor. Agradeço também, ao professor Pledson por todo incentivo. Aos meus colegas e amigos do PPGE, pela troca de conhecimentos e por toda ajuda concedida durante o doutorado, em especial, Bruna, Josimar, César, Eduardo, Anabeth, Isaac, Cristiany, João e Valéria.

À CAPES pelo apoio financeiro.

RESUMO

O objetivo da análise estatística de forma é estudar a forma dos objetos com base em pontos de referência, chamados de marcos. O caso mais comum são as formas planas em que coordenadas bidimensionais são estudadas. Existem vários tipos de sistemas para se trabalhar com esse tipo de dados, um deles é a pré-forma. A matriz de configuração é transformada em um vetor complexo, no qual os efeitos de locação e escala são removidos. Alguns métodos de aprendizado supervisionado para as pré-formas são propostos neste trabalho. Os métodos de aprendizado supervisionado são usados para classificar um objeto com base em suas informações de rótulo. O método principal é baseado em um algoritmo de máquina de vetores de suporte. Outros métodos foram baseados na estimação de densidade, classificador de Bayes, kernel *k*-médias, *hill-climbing*, por fim, uma combinação entre os classificadores utilizando o método *ensemble*. Experimentos de simulações e análises de dados reais, indicam que o algoritmo de máquina de vetores de suporte supera os outros métodos considerados quando os dados são pouco concentrados. Adicionalmente, um estudo sobre bondade de ajuste foi realizado para a distribuição Watson complexa e Bingham complexa com a utilização dos testes da razão de verossimilhanças, Kolmogorov-Smirnov, Cramér-von Mises e Anderson-Darling. Os testes foram realizados com a utilização do método *bootstrap* em uma transformação obtida por meio da aproximação dessas distribuições para normal multivariada complexa singular. Experimentos de simulações foram feitos e os testes obtiveram bons resultados, em relação ao tamanho e poder do teste. Por fim, os resultados foram ilustrados na análise de dados reais.

Palavras-chave: Aprendizado supervisionado. Bingham complexa. Pré-formas. Teste de bondade de ajuste. Watson complexa.

ABSTRACT

The aim of statistical shape analysis is to study the shape of the objects based on their landmarks. The most common case are the planar shapes where two dimensional coordinates are studied. There are several types of systems to work with this kind of data, one of them is the preshape. The configuration matrix is transformed into a complex vector, in which the location and scale effects are removed. Some methods of supervised learning for preshapes are considered in this work. Supervised learning methods are used to classify an object based on its label information. The main method is based on a support vector machine algorithm. Other methods are based on density estimation, Bayes classifier, kernel k-means, hill-climbing and a combination of classifiers using the ensemble method. Simulations experiments and real data analysis indicate that the support vector machine algorithm outperform the other considered methods, when the data are low concentrated. In addition, a goodness-of-fit study was proposed for the complex Watson and complex Bingham distribution using the likelihood ratio, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests. The tests were performed using the bootstrap method in a transformation obtained by approximating these distributions to singular complex multivariate normal. Simulation experiments were done and the tests obtained good results, such as the size and power of the test. Finally, the results were illustrated in the analysis of real data set.

Keywords: Supervised learning. Complex Bingham. Preshapes. Goodness-of-fit test. Complex Watson.

LISTA DE FIGURAS

Figura 1 – Representação do objeto e os marcos matemáticos.	32
Figura 2 – Representação gráfica de um crânio de uma gorila fêmea com 8 marcos anatômicos.	35
Figura 3 – Esquema ilustrativo das relações entre as distâncias d_F , d_P e ρ . Fonte: Dryden e Mardia (2016, p. 78).	36
Figura 4 – Pré-formas dos crânios de 29 gorilas fêmea e a forma média.	38
Figura 5 – Visão geométrica das coordenadas tangentes na esfera real. Fonte: Dryden e Mardia (2016, p. 89).	40
Figura 6 – Dados linearmente separáveis com um hiperplano separador.	49
Figura 7 – Dados linearmente separáveis com 5 hiperplanos separadores.	50
Figura 8 – Dados linearmente separáveis com o hiperplano de margem máxima.	50
Figura 9 – (a) Encontram-se pontos entre as margens e corretamente classificados; (b) Encontram-se pontos do lado incorreto, mas entre as margens; (c) Encontram-se pontos do lado incorreto e fora das margens.	55
Figura 10 – (a) visualização dos dados; (b) separação dos grupos por um hiperplano.	58
Figura 11 – Dados no espaço de características com o hiperplano separador.	59
Figura 12 – Histograma de dados simulados sobreposto por estimativas utilizando estimador ingênuo com três valores diferentes de h_1	72
Figura 13 – Esquema da validação cruzada utilizando o método $K - fold$ com $K = 5$. Fonte: Hastie, Tibshirani e Friedman (2001, p. 242)	88
Figura 14 – Formas médias para os dois grupos.	90
Figura 15 – Marcos nas coordenadas procrustes dos crânios de gorilas.	94
Figura 16 – Marcos nas coordenadas procrustes dos crânios de orangotangos.	95
Figura 17 – Marcos nas coordenadas procrustes para as vértebras de ratos.	95
Figura 18 – Marcos nas coordenadas procrustes da ressonância magnética do cérebro de pacientes com e sem esquizofrenia.	96
Figura 19 – Representação da forma média: (a) Crânio do gorila; (b) Crânio do orangotango; (c) Vértebra do rato; (d) Ressonância magnética do cérebro.	97
Figura 20 – Forma média com 3 marcos.	108
Figura 21 – Forma média com 3 marcos no espaço das pré-formas.	109

Figura 22 – Simulação no espaço das pré-formas com $n = 50$, forma média $\boldsymbol{\mu}$ da equação (4.9) e utilizando $\kappa = 10, 50, 100, 1000$	110
Figura 23 – Representação das amostras de uma distribuição normal multivariada singular.	113
Figura 24 – Histograma das amostras de uma distribuição normal multivariada singular utilizando a transformação (4.14)	114
Figura 25 – Análise gráfica utilizando a aproximação (4.22), utilizando dados da distribuição Watson complexa com $\kappa = 100$ e $n = 200$. Gráfico (a): Histograma; (b): Quantil-Quantil.	123
Figura 26 – Curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\kappa = 100$ e $n = 20, 40, 200$	132
Figura 27 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\kappa = 100$ e $n = 20, 40, 100, 200$	133
Figura 28 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\kappa = 1000$ e $n = 20, 40, 100, 200$	134
Figura 29 – Marcos nas coordenadas procrustes dos crânios de chimpanzés.	135
Figura 30 – Marcos nas coordenadas procrustes de microfósseis de conchas.	136
Figura 31 – Marcos nas coordenadas procrustes dos dígitos manuscritos.	137
Figura 32 – Marcos nas coordenadas procrustes de crânios de ratos no sétimo dia.	138
Figura 33 – Amostras da distribuição Bingham complexa, considerando a forma média $\boldsymbol{\mu}$ da equação (4.9), $n = 50$ e diferentes pares de autovalores: (a) $\lambda_1 = -2, \lambda_2 = -1$; (b) $\lambda_1 = -20, \lambda_2 = -10$; (c) $\lambda_1 = -100, \lambda_2 = -50$; (d) $\lambda_1 = -1000, \lambda_2 = -500$	147
Figura 34 – Análise gráfica utilizando a aproximação (5.2), utilizando dados da distribuição Bingham complexa com $k = 6$, $\mathbf{A} = \text{diag}(-400, -300, -200, -100, 0)$ e $n = 200$. Gráfico (a): Histograma; (b): Quantil-Quantil.	150
Figura 35 – Curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\sigma = 100$ e $n = 20, 40, 200$	156
Figura 36 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\sigma = 100$ e $n = 20, 40, 200$	157

Figura 37 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação $\boldsymbol{\varepsilon}$ com 7 valores, $\sigma = 1000$ e $n = 20, 40, 200$ 158

LISTA DE TABELAS

Tabela 1 – Exemplos de algumas funções do tipo kernel.	61
Tabela 2 – Exemplos de alguns kernels complexos.	66
Tabela 3 – Exemplos de algumas funções kernel para densidades.	73
Tabela 4 – Exemplos de alguns kernels complexos.	79
Tabela 5 – Taxas de erro para a simulação, considerando diferentes configurações de variabilidade.	93
Tabela 6 – Taxas de erro para cada classificador em relação aos 4 banco de dados, utilizando o método <i>leave-one-out</i>	99
Tabela 7 – Taxas de erro para cada classificador em relação aos dados de vértebras e macacos, utilizando o método <i>leave-one-out</i>	100
Tabela 8 – Probabilidade de aceitação ($1/M$) utilizando a distribuição CGAC para gerar amostras para a distribuição Watson complexa, utilizando vários valores do parâmetro κ e da dimensão do espaço de configuração k , para o vetor de médias $\boldsymbol{\mu} = (0, \dots, i)^\top$ com $k - 2$ zeros.	107
Tabela 9 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível de significância de 5%.	124
Tabela 10 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, para valores de $\kappa = 10, 50, 5000$, tamanhos amostrais $n = 20, 30, 40, 50, 100, 200$ e utilizando $\alpha = 1\%, 5\%, 10\%$	128
Tabela 11 – Resumo das formas médias utilizadas na análise de poder com o respectivo valor de contaminação.	130
Tabela 12 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível de significância de 5% e 7 valores de contaminação.	131
Tabela 13 – Resultados para os testes de adequação da distribuição Watson complexa utilizando os testes TRV, TKS, TCM e TAD.	139

Tabela 14 – Resultados para os testes de adequação para os dados dos crânios de $n = 18$ ratos e $k = 8$ marcos através de 8 momentos de crescimento utilizando os testes TRV, TKS, TCM e TAD.	139
Tabela 15 – Probabilidade de aceitação ($1/M$) utilizando a distribuição CGAC para gerar amostras da distribuição Bingham complexa, usando várias combinações de autovalores e da dimensão k , para uma rotação fixa.	144
Tabela 16 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível de significância de 5%.	151
Tabela 17 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , considerando a matriz $\mathbf{A} = \sigma \times \text{diag}(-4, -3, -2, -1, 0)$, para $\sigma = 10, 50, 5000$, $n = 20, 30, 40, 50, 100, 200$ e os níveis de significância de $\alpha = 1\%, 5\%, 10\%$	154
Tabela 18 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\sigma = 100, n = 20, 30, 40, 50, 100, 200$, nível de significância de 5% e 7 valores de contaminação.	155
Tabela 19 – Resultados para os testes de adequação da distribuição Bingham complexa utilizando os testes TRV, TKS, TCM, TAD.	159
Tabela 20 – Resultados para os testes de adequação da distribuição Bingham complexa para os dados dos crânios de 18 ratos e 8 marcos através de 8 momentos de crescimento utilizando os testes: TRV, TKS, TCM e TAD.	160
Tabela 21 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.	170
Tabela 22 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.	171

Tabela 23 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.	172
Tabela 24 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 5%.	173
Tabela 25 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.	174
Tabela 26 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	175
Tabela 27 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	176
Tabela 28 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	177
Tabela 29 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	178
Tabela 30 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	179
Tabela 31 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Watson complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	180

Tabela 32 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 1000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	181
Tabela 33 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 5000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	182
Tabela 34 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	183
Tabela 35 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	184
Tabela 36 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	185
Tabela 37 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	186
Tabela 38 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	187
Tabela 39 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 1000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	188
Tabela 40 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 5000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	189

Tabela 41 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	190
Tabela 42 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	191
Tabela 43 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	192
Tabela 44 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	193
Tabela 45 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	194
Tabela 46 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	195
Tabela 47 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 1000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	196
Tabela 48 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 5000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	197
Tabela 49 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.	199
Tabela 50 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.	200

Tabela 51 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.201	
Tabela 52 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 5%.202	
Tabela 53 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.203	
Tabela 54 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	204
Tabela 55 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	205
Tabela 56 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	206
Tabela 57 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	207
Tabela 58 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	208
Tabela 59 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	209

Tabela 60 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	210
Tabela 61 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 1000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	211
Tabela 62 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 5000$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.	212
Tabela 63 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	213
Tabela 64 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.	214
Tabela 65 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	215
Tabela 66 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	216
Tabela 67 – Taxas de rejeição de \mathbf{H}_0 : os dados possuem distribuição Bingham complexa, quando \mathbf{H}_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	217

Tabela 68 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	218
Tabela 69 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 1000$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	219
Tabela 70 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 5000$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.	220
Tabela 71 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.	221
Tabela 72 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	222
Tabela 73 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	223
Tabela 74 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	224
Tabela 75 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	225

Tabela 76 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	226
Tabela 77 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 100$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	227
Tabela 78 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 1000$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.	228
Tabela 79 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método <i>bootstrap</i> , para $\sigma = 5000$, $n = 20, 30, 40, 50, 100, 200$, utilizando o nível nominal de 10% e 7 valores de contaminação.	229

LISTA DE SÍMBOLOS

k	Número de <i>landmarks</i>
m	Número da dimensão
$\tilde{\mathbf{Y}}$	Matriz de configuração
\mathbf{z}^0	Configuração complexa para os <i>landmarks</i>
\mathbf{H}^F	Matrix completa de Helmert
\mathbf{H}	Submatriz de Helmert
$\boldsymbol{\omega}$	Configuração Helmertizada
$\ \cdot\ $	Norma do vetor
$\mathbf{z} = \boldsymbol{\omega} / \ \boldsymbol{\omega}\ $	Pré-forma
$\mathbb{C}S^{k-2}$	Espaço da pré-forma de dimensão $k - 1$
$*$	Operador trasposto conjugado
d_F	Distância procrustes completa
d_p	Distância procrustes parcial
ρ	Distância geodésica
$\boldsymbol{\mu}$	Forma média
$\hat{\boldsymbol{\mu}}$	Forma média amostral
$\mathbf{S} = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^*$	Matriz soma produto das pré-formas
\mathbf{I}_p	Matriz identidade de dimensão $p \times p$
$\boldsymbol{\Lambda}$	Matriz diagonal de autovalores da decomposição espectral
$\lambda_1, \lambda_2, \dots, \lambda_p$	Autovalores da decomposição espectral
$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$	Autovetores da decomposição espectral
\mathbf{U}	Matriz ortogonal de autovetores da decomposição espectral
\mathbf{v}_j	Coordenadas tangentes
$Arg(\cdot)$	Argumento de um número complexo
$Re(\cdot)$	Operador que extrai a parte real de um número ou vetor

$\text{Im}(\cdot)$	Operador que extrai a parte imaginária de um número ou vetor
$\text{cvec}(\cdot)$	Operador que empilha as partes real e imaginária de um vetor complexo
\mathbf{X}	Vetor aleatório
\mathbf{x}	Vetor observado de uma variável aleatória \mathbf{X}
g	Número de grupos
$f(\cdot)$	Densidade de uma variável aleatória
$\delta_j(\mathbf{x})$	Classificador de Bayes
\mathbf{Z}	Vetor aleatório complexo
\mathbf{z}	Vetor observado de um vetor aleatório \mathbf{Z}
$h(\mathbf{x}) = \mathbf{w}_h^\top \mathbf{x} + b_h$	Hiperplano separador para o caso real
\mathbf{w}_h	Vetor de coeficientes do hiperplano
b_h	Escalar do hiperplano real
$y_i \in \{+1, -1\}$	Rótulo dos dados de treinamento
\mathbf{w}	Novo vetor de coeficientes do hiperplano
b	Novo escalar do hiperplano
d_E	Distância euclidiana
$d(\mathbf{x}_i)$	Distância do ponto \mathbf{x}_i ao hiperplano $H(\mathbf{x})$
\mathbf{L}_p	Função primal de Lagrange
α_i	Multiplicador de Lagrange
n_{vs}	Número de vetores de suporte
ξ_i	Variável folga do hiperplano
C	Constante de penalização
η_i	Multiplicador de Lagrange
$De(\mathbf{x})$	Função de classificação da SVM
$\text{sign}(\cdot)$	Função sinal: retorna o sinal de um número real
$\phi(\mathbf{x})$	Mapeamento kernel

$K(\mathbf{x}, \mathbf{x}')$	Função kernel entre \mathbf{x} e \mathbf{x}'
$\langle \cdot, \cdot \rangle$	Produto interno
$h_{\mathbb{C}}(\mathbf{z}) = \mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}) + b$	Hiperplano complexo
$\phi_{\mathbb{C}}(\cdot)$	Mapeamento complexo
$y_i^r \in \{+1, -1\}$	Rótulo da amostra da parte real
$y_i^i \in \{+1, -1\}$	Rótulo da amostra da parte imaginária
$*$	Conjugado de um vetor complexo
$K_{\mathbb{C}}(\mathbf{z}, \mathbf{z}')$	Kernel complexo entre \mathbf{z} e \mathbf{z}'
$K_{\mathbb{C}}^r(\mathbf{z}, \mathbf{z}')$	$2\text{Re}(K_{\mathbb{C}}(\mathbf{z}, \mathbf{z}'))$
h_1	Suavizador da estimação de densidade
d_m^2	Quadrado da distância de Mahalanobis
F_H	Estatística de Hotelling
F_J	Estatística de James
F_G	Estatística de Goodall
F_{G_2}	Estatística de Goodall para g populações
λ_{\min}	Estatística Lambda para g populações
Λ_{Wilks}	Estatística de Wilks
Λ_{Roy}	Estatística de Roy
Λ_{LH}	Estatística de Lawley-Hotelling
Λ_{Pillai}	Estatística de Pillai
h_2	Suavizador do kernel procrustes

SUMÁRIO

1	INTRODUÇÃO	27
1.1	OBJETIVOS	28
1.1.1	Objetivos Específicos	28
1.1.2	Contribuições	28
1.2	ORGANIZAÇÃO DA TESE	29
1.3	SUPORTE COMPUTACIONAL	30
2	ANÁLISE ESTATÍSTICA DE FORMA	31
2.1	INTRODUÇÃO	31
2.2	MARCOS	31
2.3	SISTEMAS DE COORDENADAS	32
2.3.1	Espaço de Configuração	33
2.3.2	Distâncias entre Objetos	35
2.3.3	Coordenadas Tangentes	39
3	MÉTODOS DE CLASSIFICAÇÃO	42
3.1	INTRODUÇÃO	42
3.2	CLASSIFICADOR DE BAYES	42
3.2.1	Máxima Verossimilhança	43
3.2.2	Discriminante - Bingham Complexa	44
3.2.2.1	Inferência: Bingham Complexa	45
3.2.3	Discriminante - Watson Complexa	46
3.2.3.1	Inferência: Watson Complexa	47
3.3	<i>SUPPORT VECTOR MACHINE</i>	48
3.3.1	Dados Linearmente Separáveis	49
3.3.2	SVM Lineares	50
3.3.3	SVM Com Erros	54
3.3.4	SVM Não-Linear	57
3.3.4.1	Kernels	58
3.4	SVM - NÚMEROS COMPLEXOS	63
3.4.1	Diferenciação Complexa	63
3.4.1.1	Cálculo de <i>Wirtinger</i>	64
3.4.2	SVM Complexa	64

3.4.3	SVM com Pesos	68
3.5	ESTIMAÇÃO DE DENSIDADE	71
3.5.1	O Estimador Ingênuo	72
3.5.2	O Estimador Kernel	73
3.5.3	O Estimador Kernel Multivariado	73
3.6	<i>HILL-CLIMBING</i>	75
3.6.1	Testes para Duas Amostras	76
3.6.1.1	Estatística de Hotelling	76
3.6.1.2	Estatística de James	76
3.6.1.3	Estatística de Goodall	77
3.6.2	Testes para Várias Populações	77
3.6.2.1	Extensão da Estatística Goodall	77
3.6.2.2	Estatística Lambda	78
3.6.2.3	Estatísticas de Teste no Plano Real	78
3.7	MÉTODOS DE AGRUPAMENTO	81
3.7.1	K-médias	81
3.7.1.1	Kernel k-médias	83
3.7.1.2	Classificação Por Meio de Métodos de Agrupamento	85
3.8	MÉTODOS <i>ENSEMBLE</i>	85
3.8.1	Combinação de Classificadores	86
3.8.1.1	Maioria dos Votos	86
3.8.1.2	Voto de Pluralidade	87
3.8.1.3	Votação Ponderada	87
3.9	VALIDAÇÃO CRUZADA	88
3.10	SIMULAÇÃO	89
3.11	APLICAÇÃO EM DADOS REAIS	94
3.12	CONCLUSÃO	99
4	BONDADE DE AJUSTE PARA A DISTRIBUIÇÃO WATSON COM- PLEXA	102
4.1	INTRODUÇÃO	102
4.2	DISTRIBUIÇÃO WATSON COMPLEXA	102
4.3	SIMULADOR DA DISTRIBUIÇÃO WATSON COMPLEXA	103
4.4	DISTRIBUIÇÃO NORMAL MULTIVARIADA	111

4.4.1	Distribuição Normal Multivariada Singular	112
4.4.2	Distribuição Normal Multivariada Complexa	114
4.5	TESTES DE HIPÓTESES	115
4.5.1	Teste da Razão de Verossimilhanças	115
4.5.2	Teste de Kolmogorov	116
4.5.3	Teste de Cramér-von Mises	117
4.5.4	Teste de Anderson-Darling	117
4.5.5	Método <i>Bootstrap</i>	118
4.6	APROXIMAÇÃO PARA A WATSON COMPLEXA	119
4.6.1	Avaliação da Aproximação para a Distribuição Qui-quadrado - Watson Complexa	122
4.6.2	Distribuição de Dirichlet	125
4.7	AVALIAÇÃO NUMÉRICA	128
4.7.1	Tamanho do Teste	128
4.7.2	Poder do Teste	129
4.8	APLICAÇÃO EM DADOS REAIS	135
4.9	CONCLUSÃO	140
5	BONDADE DE AJUSTE PARA A DISTRIBUIÇÃO BINGHAM COMPLEXA	141
5.1	INTRODUÇÃO	141
5.2	DISTRIBUIÇÃO BINGHAM COMPLEXA	141
5.3	SIMULADOR DA DISTRIBUIÇÃO BINGHAM COMPLEXA	142
5.4	ORTOGONALIZAÇÃO DE GRAM-SCHMIDT	145
5.4.1	Simulador da Distribuição Bingham Complexa via Gram-Schmidt . . .	146
5.5	APROXIMAÇÃO PARA A BINGHAM COMPLEXA	148
5.5.1	Avaliação da Aproximação da Distribuição Qui-quadrado - Bingham Complexa	149
5.6	AVALIAÇÃO NUMÉRICA - SIMULAÇÃO	153
5.6.1	Tamanho do Teste	153
5.6.2	Poder do Teste	154
5.7	APLICAÇÃO EM DADOS REAIS	159
5.8	CONCLUSÃO	161
6	CONSIDERAÇÕES FINAIS	162

REFERÊNCIAS	164
APÊNDICE A – SIMULAÇÕES DOS TESTES DE BONDADDE DE AJUSTE	
- WATSON COMPLEXA	169
APÊNDICE B – SIMULAÇÕES DOS TESTES DE BONDADDE DE AJUSTE	
- BINGHAM COMPLEXA	198

1 INTRODUÇÃO

O campo da análise estatística de forma foi introduzido por Kendall (1977). Todavia, a formulação matemática e alguns conceitos básicos foram apresentados em Kendall (1984). Bookstein (1984), Bookstein (1986) também desenvolveu uma teoria para análise estatística de forma e, mais tarde, ambos os pesquisadores definiram as conexões entre os trabalhos. Existem alguns livros sobre a análise estatística de forma. O livro de Dryden e Mardia (2016) explica várias questões relacionadas aos aspectos estatísticos, análise de dados de formas de objetos e também apresenta um bom histórico sobre os métodos. Outros livros que apresentam aspectos mais teóricos são o livro de Small (1996) e Kendall *et al.* (2009).

Atualmente, existem muitas imagens disponíveis que podem ser obtidas por meio de: câmera fotográfica, satélite, raio-X, etc. Essas imagens podem ser extraídas em vários campos, como na Biologia, Zoologia, Medicina, Oceanografia, entre outras. Uma rotina na análise estatística de forma é colocar marcos nessas imagens. As imagens podem estar em duas ou três dimensões. O foco deste trabalho está no caso bidimensional, que é chamado formas planas. A análise estatística baseada em pontos de referência é útil para várias questões, tais como: regiões de confiança, testes de hipóteses, análise de agrupamento e classificação. Os métodos relacionados à classificação estão incorporados na área de aprendizado de máquina. O aprendizado de máquina (*machine learning*) é uma vertente da inteligência artificial que proporciona identificar padrões e tomar decisões com o mínimo de intervenção humana. Esse tema vem crescendo nos últimos anos devido aos enormes conjuntos de dados (*Big Data*) e o interesse em extrair informações relevantes para a classificação. No aprendizado supervisionado, os grupos (ou classes) são conhecidos *a priori* e é necessário fornecer exemplos para o treinamento. Diferentemente, no aprendizado não supervisionado, o algoritmo tenta reconhecer as classes através de padrões existentes com um determinado critério sem receber nenhuma informação das classes (HASTIE; TIBSHIRANI; FRIEDMAN, 2001, p. xi).

Dryden e Mardia (2016) apresentam uma seção sobre classificação para formas planas. Para expandir essa área de pesquisa, este trabalho apresenta vários classificadores que são apropriados para formas planas. Na análise estatística de forma, os vetores estão compreendidos no espaço dos números complexos. Os métodos propostos são desenvolvidos para lidar com essas questões. Assim, o presente estudo analisa dados no espaço da pré-forma, isto é, dados multidimensionais em uma hipersfera complexa de raio um. Com o intuito de identificar padrões para a classificação em duas ou mais classes no espaço da pré-forma, foram propostos vários

métodos para esse fim, dentre eles estão a máquina de suporte vetorial com pesos baseados em distâncias, classificador de Bayes, método kernel (estimação de densidade), procrustes kernel k-médias e o *Hill-Climbing*. Ademais, foi proposta uma combinação desses classificadores, com o intuito de melhorar os resultados.

Outro ponto importante na análise estatística de forma é a carência em métodos de bondade de ajuste em distribuições. A principal dificuldade é pelo fato de que distribuições multivariadas complexas são de difícil manipulação algébrica e computacional. As duas principais distribuições existentes são: a distribuição Bingham complexa proposta por Kent (1994) e a distribuição Watson complexa, que é um caso especial da Bingham complexa. Existem vários testes de bondade de ajuste, dentre eles, pode-se citar: Kolmogorov-Smirnov, Cramér-von Mises e Anderson-Darling que se baseiam na distribuição empírica dos dados e na função de distribuição.

1.1 OBJETIVOS

Os objetivos gerais desse trabalho são: utilizar classificadores da literatura e os principais testes de bondade de ajuste em distribuições no contexto da análise estatística de forma.

1.1.1 Objetivos Específicos

1. Realizar uma revisão da literatura sobre a análise estatística de forma;
2. Utilizar métodos usuais de classificação e propor modificações;
3. Utilizar testes de bondade de ajuste para as distribuições Watson e Bingham complexa;
4. Realizar simulações para verificar o desempenho dos métodos de classificação e bondade de ajuste.

1.1.2 Contribuições

As principais contribuições do presente trabalho são:

1. São propostos os classificadores: máquina de suporte vetorial com pesos baseados em distâncias, Bayes com a distribuição Watson complexa, Bayes com estimação de densidade não paramétrica, procrustes kernel k-médias com protótipo e combinações entre classificadores através do método *ensemble*;
2. Um simulador de números pseudo-aleatórios para a distribuição Watson complexa;

3. Dois simuladores de números pseudo-aleatórios para a distribuição Bingham complexa;
4. Avaliações numéricas sobre a aproximação da distribuição Bingham e Watson complexa para a distribuição normal multivariada complexa singular;
5. Testes de bondade de ajuste por meio do método *bootstrap* para a distribuição Bingham e Watson complexa.

1.2 ORGANIZAÇÃO DA TESE

Além deste capítulo de introdução, esta tese é composta por mais cinco capítulos. No Capítulo 2 são apresentados conceitos sobre análise estatística de forma, como por exemplo: os tipos de marcos, as coordenadas de Kendall, distâncias entre dois objetos, o conceito de forma média e as coordenadas tangentes. No Capítulo 3 são definidos alguns métodos de classificação com seus aspectos matemáticos e vários resultados de simulações e aplicações em dados reais. Ademais, é proposto um novo classificador baseado em um sistema de pesos que utiliza a máquina de suporte vetorial, métodos de agrupamento para classificação e uma combinação entre os classificadores. No Capítulo 4 é proposto um simulador para a distribuição Watson complexa. Também é descrita uma relação entre a distribuição normal multivariada singular e a distribuição normal multivariada, e suas extensões para o caso complexo. Além disso, é avaliada uma aproximação da distribuição Watson complexa pela normal multivariada complexa singular, no qual essa aproximação é utilizada para encontrar uma distribuição aproximada para uma forma quadrática. Avaliou-se também, a qualidade dessa aproximação por meio de alguns testes de bondade de ajuste. Foi proposto uma transformação dessa forma quadrática e testes com o método *bootstrap* para verificar a bondade de ajuste da distribuição Watson complexa quando os parâmetros são estimados. Além disto, foram realizados estudos de simulações e aplicações em dados reais. No Capítulo 5 é proposto um novo simulador para a distribuição Bingham complexa via ortogonalização de Gram-Schmidt, é avaliada uma aproximação para normal multivariada complexa singular e a distribuição aproximada de uma forma quadrática que é utilizada averiguar a bondade de ajuste da distribuição Bingham complexa. Ademais, também é utilizada uma transformação nessa forma quadrática e estudos de simulações e aplicações em dados reais para a bondade de ajuste da distribuição Bingham complexa. Por fim, no Capítulo 6, são apresentadas conclusões obtidas do presente trabalho e direções para trabalhos futuros.

1.3 SUPORTE COMPUTACIONAL

Todos os gráficos e resultados no presente trabalho foram obtidos na plataforma computacional R, em sua versão 3.5.3. (R Core Team, 2019). O pacote principal do R utilizado no presente trabalho, foi o pacote `shapes`, no qual contém as bases de dados e todas as ferramentas envolvidas na análise de forma (DRYDEN, 2018).

2 ANÁLISE ESTATÍSTICA DE FORMA

2.1 INTRODUÇÃO

A análise de forma (do inglês *Shape Analysis*) é uma área da estatística responsável em estudar a forma geométrica dos objetos. Decisões sobre esses objetos são tomadas através da análise geométrica do tamanho e forma. Por exemplo, imagens de crânios de gorilas de uma determinada espécie em dois momentos diferentes, em que o interesse é verificar se houve alguma diferença entre a forma média nesses dois momentos, ou ainda, diferença entre os tamanhos desses crânios. Poderia-se comparar também, diferenças entre crânios de gorilas em relação ao sexo. Dessa forma, pode-se ver uma variedade de aplicações como, por exemplo, diagnóstico de doenças, reconhecimento facial e identificação de proteínas (DRYDEN; MARDIA, 2016, p. 1). A seguir são apresentadas definições extraídas do livro de Dryden e Mardia (2016, p. 1-3, 64) para o entendimento da análise de forma.

Definição 1. A **forma** é toda a informação geométrica que permanece quando os efeitos de locação, rotação e escala são removidos do objeto.

Pode-se dizer que a forma de um objeto é invariante sob as transformações de translação, escala e rotação.

Definição 2. O **tamanho e forma**, é quando são retirados os efeitos de locação e rotação.

Nesse caso, pode-se concluir o tamanho e forma de um objeto é invariante sob as transformações de translação e rotação. A análise de tamanho e forma é de grande interesse para várias áreas, como na biologia, química, medicina, arqueologia, geografia, etc.

Definição 3. A **pré-forma**, é quando são removidos os efeitos de locação e escala.

Diante do que foi exposto, será apresentada uma breve descrição sobre os tipos de marcos na Seção 2.2 e na Seção 2.3 os aspectos matemáticos da análise de forma como: sistemas de coordenadas, espaço de configuração, distância entre objetos e as coordenadas tangentes.

2.2 MARCOS

A forma de um objeto pode ser descrita por um conjunto finito de pontos que são chamados de marcos (do inglês *landmarks*). Segundo Dryden e Mardia (2016, p. 3), um marco é um ponto de correspondência em que cada objeto coincide entre e dentro das populações.

Basicamente, existem três tipos de marcos (DRYDEN; MARDIA, 2016, p. 3-4)

- **Marcos Anatômicos:** Pontos escolhidos entre os objetos por um especialista contendo características biológicas significativas, por exemplo, o encontro entre duas suturas em um crânio;
- **Marcos Matemáticos:** Pontos atribuídos em objetos com alguma propriedade matemática ou geométrica, por exemplo: um ponto de alta curvatura ou um ponto extremo;
- **Pseudo-Marcos:** Pontos designados nos objetos, localizados no contorno ou entre os marcos matemáticos e/ou anatômicos.

Como ilustração, pode-se ver na Figura 1 a representação de marcos matemáticos aplicados em uma imagem de uma borboleta (essa imagem encontra-se em Grátispng (2019)). Na Figura 1 (a), temos a imagem de um objeto, nesse caso uma borboleta; na Figura 1 (b) temos o objeto sobreposto pelos marcos matemáticos; na Figura 1 (c) os 20 marcos que representam o objeto matematicamente.

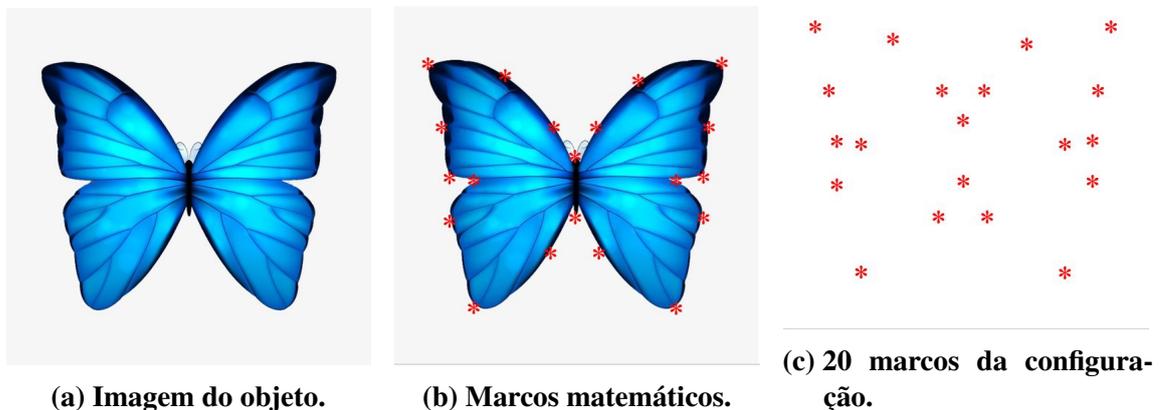


Figura 1 – Representação do objeto e os marcos matemáticos.

2.3 SISTEMAS DE COORDENADAS

Para se obter a forma é preciso especificar um sistema de coordenadas e definir matematicamente várias quantidades e operações de interesse. Existem vários sistemas de coordenadas, dentre eles, destacam-se as coordenadas de Kendall propostas por Kendall (1984) e as coordenadas de Bookstein propostas por Bookstein (1984), Bookstein (1986). Existe uma certa relação entre essas duas coordenadas. Entretanto, no presente estudo, será utilizado o sistema de coordenadas de Kendall e as definições apresentadas a seguir serão fundamentais para o seu entendimento.

2.3.1 Espaço de Configuração

Seja $\tilde{\mathbf{Y}}$ uma matriz de dimensão $k \times m$ contendo as coordenadas cartesianas de k marcos em m dimensões. A matriz $\tilde{\mathbf{Y}}$ é chamada de matriz de configuração e é representada por

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{y}_{11} & \cdots & \tilde{y}_{1m} \\ \vdots & \ddots & \vdots \\ \tilde{y}_{k1} & \cdots & \tilde{y}_{km} \end{pmatrix}.$$

No presente trabalho, são consideradas as situações em que $k \geq 3$ e $m = 2$, correspondendo as formas planas.

Com o intuito de remover os efeitos de locação, rotação e escala são necessárias algumas transformações na matriz $\tilde{\mathbf{Y}}$. Considerando $m = 2$, a matriz $\tilde{\mathbf{Y}}$ pode ser reescrita como um vetor complexo de dimensão $k \times 1$ da seguinte maneira

$$\mathbf{z}^0 = (\tilde{y}_{11} + i\tilde{y}_{12}, \dots, \tilde{y}_{k1} + i\tilde{y}_{k2})^\top, \quad (2.1)$$

que correspondem às coordenadas complexas para os marcos.

Inicialmente, para remover o efeito da locação é necessário multiplicar o vetor complexo \mathbf{z}^0 por uma matriz de centralização. Essa matriz é a submatriz de Helmert (diga-se, \mathbf{H}). Essa submatriz tem dimensão $(k-1) \times k$ e é formada pela matriz completa de Helmert sem a primeira linha. A matriz de Helmert completa (\mathbf{H}^F) é ortogonal, possui dimensão $k \times k$, nos quais os elementos da primeira linha são iguais a $1/\sqrt{k}$ e os elementos da $(j+1)$ -ésima linha com $j = 1, \dots, k-1$ são dados por

$$\mathbf{h}_{j+1} = (h_j, \dots, h_j, -jh_j, 0, \dots, 0)^\top, \quad h_j = -[j(j+1)]^{-1/2},$$

no qual o número de zeros na $(j+1)$ -ésima linha é igual a $k-j-1$. Assim, a submatriz é dada por

$$\mathbf{H} = (\mathbf{h}_2, \dots, \mathbf{h}_k)^\top.$$

Por exemplo, se $k = 4$, a matriz de Helmert é dada por

$$\mathbf{H}^F = \begin{pmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} & 0 \\ -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} \end{pmatrix},$$

e a submatriz de Helmert é

$$\mathbf{H} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} & 0 \\ -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} \end{pmatrix}.$$

Portanto, para remover o efeito da locação, basta multiplicar o vetor complexo \mathbf{z}^0 dado na equação (2.1) pela submatriz de Helmert,

$$\boldsymbol{\omega} = \mathbf{H}\mathbf{z}^0, \quad (2.2)$$

em que $\boldsymbol{\omega}$ representa a configuração \mathbf{z}^0 sem o efeito da locação. Veja que, $\boldsymbol{\omega}$ tem dimensão $(k-1)$, ou seja, quando o efeito da locação é removido, há uma perda de uma dimensão. Para que a configuração centrada tenha a mesma dimensão da configuração original, basta multiplicar a equação (2.2) por \mathbf{H}^\top .

Para remover o efeito da escala, basta dividir a configuração Helmertizada dada na equação (2.2) pela sua norma, assim

$$\mathbf{z} = \frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} = \frac{\boldsymbol{\omega}}{\sqrt{\boldsymbol{\omega}^* \boldsymbol{\omega}}} = \frac{\mathbf{H}\mathbf{z}^0}{\sqrt{(\mathbf{H}\mathbf{z}^0)^* \mathbf{H}\mathbf{z}^0}}, \quad (2.3)$$

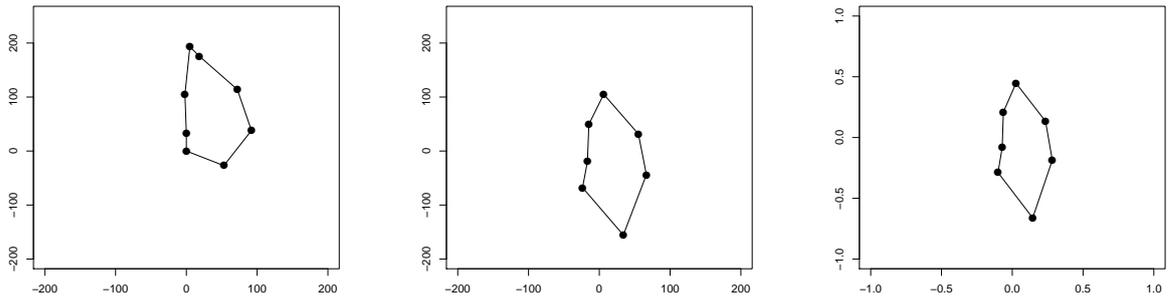
no qual $*$ é o operador transposto conjugado e $\|\cdot\|$ é a norma complexa do vetor. Além disso, \mathbf{z} é chamado de pré-forma de \mathbf{z}_0 (KENDALL, 1984). Note que \mathbf{z} é uma forma que conserva a rotação. Formalmente, o espaço da pré-forma é definida como o espaço de todos os vetores complexos possíveis em que cada vetor tem dimensão $(k-1)$ e não possui os efeitos da locação e escala. Desse modo, o espaço da pré-forma é uma hiperesfera complexa de dimensão $(k-1)$, definida por

$$\mathbb{C}S^{k-2} = \left\{ \mathbf{z} : \|\mathbf{z}\| = 1, \mathbf{z} \in \mathbb{C}^{k-1} \right\}, \quad (2.4)$$

no qual \mathbb{C}^{k-1} é o espaço dos números complexos de dimensão $(k-1)$. Existe ainda o espaço da forma, que é definido como o espaço da pré-forma sem o efeito da rotação. O espaço da forma pode ser representado da seguinte maneira

$$[\mathbf{z}] = \left\{ e^{i\theta} \mathbf{z} : \theta \in [0, 2\pi) \text{ e } \mathbf{z} \in \mathbb{C}^{k-1}, \|\mathbf{z}\| = 1 \right\},$$

em que $[\mathbf{z}]$ é qualquer versão rotacionada de \mathbf{z} . Pode-se ver que existe uma variedade de maneiras para se trabalhar com marcos. Entretanto, o enfoque do trabalho são análises no espaço da pré-forma. Na Figura 2 estão a configuração original (a), Helmertizada (b) e a pré-forma (c) para os 8 marcos de um crânio de uma gorila fêmea. Para uma melhor entendimento sobre esse dado, ver seção 3.11.



(a) Configuração original. (b) Helmertizada. (c) Pré-forma.
Figura 2 – Representação gráfica de um crânio de uma gorila fêmea com 8 marcos anatômicos.

2.3.2 Distâncias entre Objetos

Esta seção introduz algumas distâncias que serão utilizadas no presente trabalho e mostra como obter a forma média de uma amostra de pré-formas.

Sejam \mathbf{z}_1 e \mathbf{z}_2 as pré-formas de $\tilde{\mathbf{Y}}_1$ e $\tilde{\mathbf{Y}}_2$, respectivamente. Considere $|\cdot|$ como sendo o módulo de um número complexo, dessa forma, as distâncias entre \mathbf{z}_1 e \mathbf{z}_2 são definidas como (DRYDEN; MARDIA, 2016, p. 69-74)

- Distância procrustes completa

$$d_F(\mathbf{z}_1, \mathbf{z}_2) = \sqrt{1 - |\mathbf{z}_1^* \mathbf{z}_2|^2}, \quad 0 \leq d_F \leq 1. \quad (2.5)$$

- Distância procrustes parcial

$$d_P(\mathbf{z}_1, \mathbf{z}_2) = \sqrt{2(1 - |\mathbf{z}_1^* \mathbf{z}_2|)}, \quad 0 \leq d_P \leq \sqrt{2}. \quad (2.6)$$

- Distância geodésica

$$\rho(\mathbf{z}_1, \mathbf{z}_2) = \arccos(|\mathbf{z}_1^* \mathbf{z}_2|), \quad 0 \leq \rho \leq \pi/2. \quad (2.7)$$

Na Figura 3, pode-se ver um esquema sobre as relações entre essas distâncias.

Com essas distâncias é possível estimar a forma média de uma amostra de objetos. Com isso, considere uma amostra $\mathbf{z}_1^0, \dots, \mathbf{z}_n^0$ de configurações complexas de tamanho n , em que cada \mathbf{z}_i^0 é definido na equação (2.1). Considere agora $\mathbf{z}_1, \dots, \mathbf{z}_n$ obtidas usando a equação (2.3) como sendo as pré-formas de $\mathbf{z}_1^0, \dots, \mathbf{z}_n^0$, respectivamente. Assim, a forma média amostral pode ser obtida através de

$$\hat{\boldsymbol{\mu}} = \arg \inf_{\boldsymbol{\mu}} \sum_{i=1}^n d^2(\mathbf{z}_i, \boldsymbol{\mu}),$$

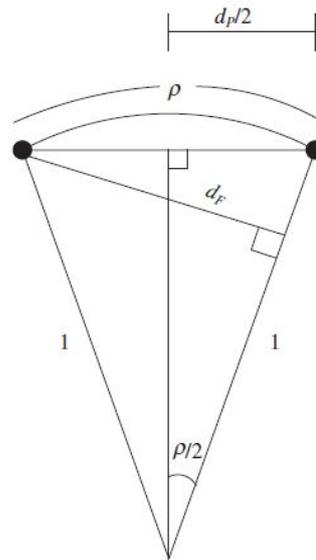


Figura 3 – Esquema ilustrativo das relações entre as distâncias d_F , d_P e ρ . Fonte: Dryden e Mardia (2016, p. 78).

em que $d(\cdot, \cdot)$ pode ser qualquer uma das três distâncias mencionadas anteriormente. Entretanto, a estimativa da forma média difere em relação as distâncias, porém não há grandes diferenças (DRYDEN; MARDIA, 2016, p. 118-119). Dessa forma, será utilizada a distância procrustes completa para estimar a forma média. Assim, é necessário minimizar

$$\begin{aligned} \sum_{i=1}^n d_F^2(\mathbf{z}_i, \boldsymbol{\mu}) &= \sum_{i=1}^n (1 - |\mathbf{z}_i^* \boldsymbol{\mu}|^2) \\ &= n - \sum_{i=1}^n \boldsymbol{\mu}^* \mathbf{z}_i \mathbf{z}_i^* \boldsymbol{\mu} = n - \boldsymbol{\mu}^* \mathbf{S} \boldsymbol{\mu}, \end{aligned} \quad (2.8)$$

em que

$$\mathbf{S} = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^*. \quad (2.9)$$

Note que para minimizar (2.8), basta maximizar $\boldsymbol{\mu}^* \mathbf{S} \boldsymbol{\mu}$. Desse modo, para resolver esse problema é preciso utilizar alguns resultados.

Definição 4. Para toda matriz \mathbf{A} complexa de dimensão $p \times p$, pode-se encontrar uma constante λ e um vetor complexo $\mathbf{u} \in \mathbb{C}^p$ não-nulo, tais que

$$\mathbf{A} \mathbf{u} = \lambda \mathbf{u},$$

em que λ é chamado de autovalor e \mathbf{u} de autovetor com $\|\mathbf{u}\| = 1$.

Para encontrar os autovalores basta resolver a equação (HORN, 1990, p. 38)

$$|\mathbf{A} - \lambda \mathbf{I}_p| = 0,$$

no qual \mathbf{I}_p é a matriz identidade de dimensão $p \times p$. Em seguida, encontrar os autovetores substituindo os autovalores na definição.

Definição 5. Uma matriz complexa \mathbf{A} , é dita ser Hermitiana se $\mathbf{A} = \mathbf{A}^*$.

Definição 6. Uma matriz Hermitiana \mathbf{A} , é definida positiva ou positiva definida, se

$$\mathbf{z}^* \mathbf{A} \mathbf{z} > 0, \forall \mathbf{z} \in \mathbb{C}^p.$$

Caso $\mathbf{z}^* \mathbf{A} \mathbf{z} \geq 0$, então a matriz é dita semidefinida.

Pode-se ver claramente que a matriz \mathbf{S} é Hermitiana, pois

$$\mathbf{S}^* = \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^* \right)^* = \sum_{i=1}^n (\mathbf{z}_i \mathbf{z}_i^*)^* = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^* = \mathbf{S},$$

e é chamada de matriz soma produto das pré-formas. Além disso, \mathbf{S} também é positiva definida, visto que

$$\mathbf{z}^* \mathbf{S} \mathbf{z} = \mathbf{z}^* \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^* \right) \mathbf{z} = \sum_{i=1}^n (\mathbf{z}^* \mathbf{z}_i) (\mathbf{z}_i^* \mathbf{z}) = \sum_{i=1}^n |\mathbf{z}^* \mathbf{z}_i|^2 > 0,$$

já que $\mathbf{z}^* \mathbf{z}_i$ é um número complexo e $\mathbf{z}_i^* \mathbf{z}$ é seu conjugado. Dessa forma, $(\mathbf{z}^* \mathbf{z}_i)(\mathbf{z}_i^* \mathbf{z})$ é um número real maior que 0.

Teorema 1 (Decomposição Espectral). Seja \mathbf{A} uma matriz Hermitiana de dimensão $p \times p$, então

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^* = \sum_{i=1}^p \lambda_i \mathbf{u}_i \mathbf{u}_i^*$$

no qual $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, são os autovalores de \mathbf{A} e \mathbf{U} é uma matriz ortogonal ($\mathbf{U}^* \mathbf{U} = \mathbf{I}$), com a i -ésima coluna sendo \mathbf{u}_i , que corresponde ao i -ésimo autovetor correspondente ao i -ésimo autovalor da matriz \mathbf{A} .

A prova do Teorema 1 pode ser vista em Mirsky (1955, p. 387). Caso a matriz \mathbf{A} também seja positiva definida, então, pode-se mostrar que todos os autovalores de \mathbf{A} são maiores que 0. De fato, se \mathbf{A} é positiva, então, $\mathbf{z}^* \mathbf{A} \mathbf{z} > 0$ (Definição 6). Substituindo \mathbf{z} pelo i -ésimo autovetor de \mathbf{A} , temos que

$$\mathbf{u}_i^* \mathbf{A} \mathbf{u}_i = \lambda_i > 0, \forall i = 1, 2, \dots, p.$$

Teorema 2. Seja \mathbf{A} uma matriz Hermitiana positiva definida, então

$$\max_{\mathbf{z} \neq 0} \mathbf{z}^* \mathbf{A} \mathbf{z} = \lambda_i, \text{ se } \mathbf{z} = \mathbf{u}_i,$$

para $i = 1, 2, \dots, p$.

Para mais detalhes sobre o Teorema 2 ver Johnson e Wichern (2007, p. 80-81). Foi visto que para encontrar a forma média, basta maximizar $\boldsymbol{\mu}^* \mathbf{S} \boldsymbol{\mu}$. Pelo Teorema 2,

$$\max_{\boldsymbol{\mu}} \boldsymbol{\mu}^* \mathbf{S} \boldsymbol{\mu} = \lambda_1, \text{ quando } \boldsymbol{\mu} = \mathbf{u}_1.$$

Portanto, o estimador da forma média é dado por

$$\hat{\boldsymbol{\mu}} = \mathbf{u}_1, \quad (2.10)$$

que representa o autovetor correspondente ao maior autovalor de \mathbf{S} (KENT, 1994). A forma média desempenha um papel muito importante na análise de forma, como em testes de hipóteses para a comparação das formas médias de dois grupos ou mais, ou ainda, o seu uso em sistemas de coordenadas. A Figura 4 possui as pré-formas dos crânios de 29 gorilas fêmea, juntamente com a forma média (2.10).

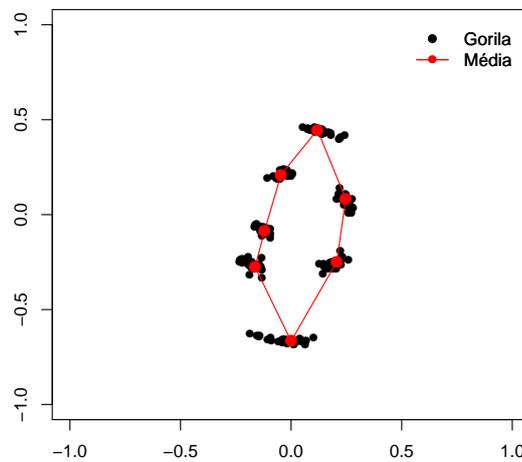


Figura 4 – Pré-formas dos crânios de 29 gorilas fêmea e a forma média.

2.3.3 Coordenadas Tangentes

As coordenadas tangentes (CT) são formadas pela projeção das formas no plano tangente na vizinhança de um polo, chamado de polo da projeção tangente. Geralmente, esse polo é a forma média dos objetos (DRYDEN; MARDIA, 2016, p. 88). Vamos considerar no estudo as CT procrustes, em que são utilizadas as pré-formas, com isso, os pontos da esfera complexa são projetados no plano tangente complexo e, ademais, os novos pontos no plano complexo são transformados em pontos no plano real. Assim, é possível utilizar as técnicas multivariadas usuais, como o discriminante linear, análise de componentes principais, testes de comparações de médias, etc.

A ideia das CT é particionar um vetor em outras duas parcelas, esse particionamento é chamado de decomposição tangente normal (LEY; VERDEBOUT, 2017, p. 38). Como ilustração, considere $\mathbf{a} = (1, 0, 0)^\top$ e $\mathbf{b} = (0, 1, 0)^\top$ dois vetores unitários. Então qualquer vetor $\mathbf{x} = (x_1, x_2, 0)^\top$ pode ser expresso como

$$\mathbf{x} = x_1 \mathbf{a} + x_2 \mathbf{b}.$$

Dessa forma, pode-se ver que um vetor pode facilmente ser particionado em duas ou mais componentes (vetores ou matrizes). Assim, a decomposição tangente normal é dada por

$$\mathbf{x} = t\xi + (1 - t^2)^{1/2} \boldsymbol{\gamma}.$$

Impondo que $\xi^\top \mathbf{x} = 0$, então para que a igualdade seja válida, $(1 - t^2)^{1/2} = \mathbf{x}^\top \boldsymbol{\gamma}$, consequentemente

$$t\xi = [\mathbf{I}_p - \boldsymbol{\gamma}\boldsymbol{\gamma}^\top] \mathbf{x},$$

e

$$\xi = \frac{[\mathbf{I}_p - \boldsymbol{\gamma}\boldsymbol{\gamma}^\top] \mathbf{x}}{\|[\mathbf{I}_p - \boldsymbol{\gamma}\boldsymbol{\gamma}^\top] \mathbf{x}\|}.$$

A matriz $\mathbf{I}_p - \boldsymbol{\gamma}\boldsymbol{\gamma}^\top$ é a matriz de projeção no espaço ortogonal a $\boldsymbol{\gamma}$ (DRYDEN; MARDIA, 2016, p. 89). Observe que essa decomposição também é válida para vetores complexos. Na Figura 5, pode-se ver uma ilustração da decomposição tangente normal fazendo $\boldsymbol{\gamma} = \boldsymbol{\mu}$.

Considere uma amostra de pré-formas $\mathbf{z}_1, \dots, \mathbf{z}_n$, assim sendo, as CT procrustes são dadas por

$$\mathbf{t}_j = e^{i\hat{\theta}_j} [\mathbf{I}_{k-1} - \boldsymbol{\mu}\boldsymbol{\mu}^*] \mathbf{z}_j, \quad j = 1, \dots, n,$$

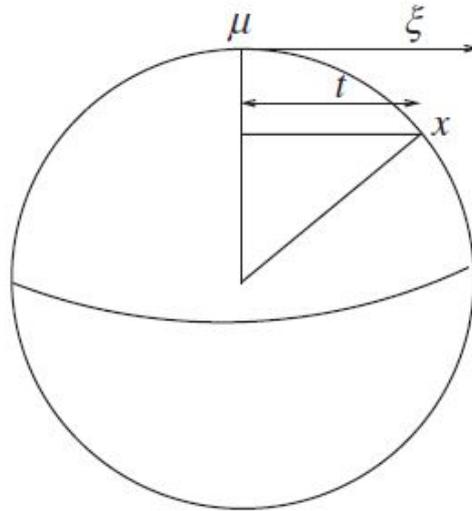


Figura 5 – Visão geométrica das coordenadas tangentes na esfera real. Fonte: Dryden e Mardia (2016, p. 89).

em que \mathbf{I}_{k-1} é a matriz identidade $(k-1) \times (k-1)$, $\hat{\theta}_j = \text{Arg}(-\boldsymbol{\mu}^* \mathbf{z}_j)$ é o valor que minimiza $\|\boldsymbol{\mu} - \mathbf{z}e^{i\theta}\|^2$ (no qual $\text{Arg}(a)$, com $a \in \mathbb{C}$, é o ângulo compreendido entre o eixo real positivo no plano complexo e a reta que une a com a origem desse plano) e $\boldsymbol{\mu}$ é a forma média (DRYDEN; MARDIA, 2016, p. 91). Geralmente, como não se conhece a forma média, basta substituí-la pela forma média amostral (2.10). Veja que essas coordenadas possuem o termo $e^{i\hat{\theta}}$, que tem como objetivo rotacionar os objetos para que eles fiquem o mais próximo possível da forma média. Com isso, melhorando a representação dos objetos no plano tangente.

Normalmente, quando se utiliza esse sistema de coordenadas, é feito um empilhamento das coordenadas real e imaginária de cada \mathbf{t}_j . Seja \mathbf{v}_j um vetor de dimensão $2k-2$, que irá armazenar as coordenadas de \mathbf{t}_j da seguinte maneira

$$\mathbf{v}_j = \text{cvec}(\mathbf{t}_j) = (\text{Re}(\mathbf{t}_j)^\top, \text{Im}(\mathbf{t}_j)^\top)^\top, \quad (2.11)$$

em que $\text{cvec}(\cdot)$ é o operador que empilha as partes real e imaginárias de um vetor complexo, transformando-o em um vetor real, $\text{Re}(\cdot)$ caracteriza a operação de extração da parte real do vetor e $\text{Im}(\cdot)$ a parte imaginária. Assim sendo, cada vetor de pré-formas \mathbf{z}_j é representado nas CT pelo vetor \mathbf{v}_j .

Um outro tipo de coordenadas na análise de formas são as coordenadas procrustes, cujo intuito é casar as formas de dois objetos independente da locação, rotação e escala. Geralmente, esse ajuste é feito para casar os objetos com a forma média amostral (2.10), ou seja, rotacionar os objetos para que eles sejam o mais próximo possível da forma média amostral. As-

sim, esse ajuste é definido como ajuste procrustes completo ou coordenadas procrustes completas de $\mathbf{z}_1, \dots, \mathbf{z}_n$ como sendo

$$\mathbf{z}_i^P = \mathbf{z}_i^* \hat{\boldsymbol{\mu}} \mathbf{z}_i, \quad i = 1, \dots, n, \quad (2.12)$$

no qual cada \mathbf{z}_i^P é dito ser o ajuste procrustes completo de \mathbf{z}_i em $\hat{\boldsymbol{\mu}}$, para mais detalhes ver Dryden e Mardia (2016, p. 24). Geralmente, essas coordenadas são utilizadas para criar gráficos sobre a população em estudo. Outro ponto importante, é que nessas coordenadas o vetor médio amostral $(1/n) \sum_{i=1}^n \mathbf{z}_i^P$ coincide com a forma média procrustes completa (DRYDEN; MARDIA, 2016, p. 24).

Existem ainda os resíduos do ajuste procrustes, que são utilizados para estudar a variabilidade da forma. Os resíduos são calculados da seguinte maneira

$$\mathbf{e}_j = \mathbf{z}_j^P - \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i^P, \quad j = 1, \dots, n.$$

3 MÉTODOS DE CLASSIFICAÇÃO

3.1 INTRODUÇÃO

O Aprendizado Estatístico (AE) (do inglês, *Statistical Learning*), desempenha um papel importante em várias áreas da ciência, finanças e indústria. O objetivo do AE é fornecer uma estrutura para a inferência, ganhar conhecimento, fazer previsões, tomar decisão ou construir modelos através de um conjunto de dados. Pode-se citar como exemplo: prever a reincidência de um câncer; estimar a quantidade de glicose no sangue de uma pessoa com diabetes; reconhecimento de voz; reconhecimento facial; reconhecimento de letras e números, etc. Na aprendizagem de máquina existem algumas categorias e as principais são: supervisionado e não-supervisionado. O aprendizado supervisionado consiste em aprender com um conjunto de dados de treinamento, ou seja, as classes (grupos ou rótulos) de um conjunto de dados são conhecidas e desejamos criar uma função para prever a classe de um dado futuro. Já no aprendizado não-supervisionado, as classes dos dados não são fornecidas e o objetivo é descrever as associações e os padrões entre o conjunto de treinamento. Uma das principais referências nessa área de aprendizado é o livro de Hastie, Tibshirani e Friedman (2001). No presente estudo, são considerados problemas com aprendizado supervisionado.

Neste Capítulo serão apresentados vários métodos de classificação dos mais diferentes tipos. Na Seção 3.2 é apresentado o classificador de Bayes com algumas distribuições no espaço das pré-formas e o método da máxima verossimilhança; já na Seção 3.3 é mostrada toda a ideia da máquina de vetores de suporte para o caso real; na Seção 3.4 é definido a máquina de vetores de suporte para o caso complexo e uma proposta de modificação na função de otimização do problema; na Seção 3.5 é exibido um método não paramétrico para estimar densidades, que se tornará um classificador com a utilização do classificador de Bayes; por sua vez, nas Seções 3.6 e 3.7 são propostos classificadores baseados em estatísticas de testes e com a utilização de métodos de agrupamento; na Seção 3.8 são exibidas três maneiras para combinar classificadores; por fim, na Seção 3.9 é descrito o método da validação cruzada e as simulações e aplicações em dados reais são descritas nas Seções 3.10 e 3.11, respectivamente.

3.2 CLASSIFICADOR DE BAYES

O classificador de Bayes, também conhecido como regra de Bayes ou *naive Bayes*, é um método de classificação probabilístico, pois é levado em conta a distribuição dos dados.

Primeiramente, deve-se escolher uma distribuição para os dados e em seguida estimar os seus parâmetros.

Seja $\mathbf{X} \in \mathbb{R}^p$ um vetor aleatório, sendo \mathbf{x} o valor observado correspondente, em que desejamos classificar um objeto dentre um dos g grupos ($g \geq 2$). Seja π_i a probabilidade *a priori* de um objeto pertencer ao grupo i ($i = 1, \dots, g$), com $\sum_{i=1}^g \pi_i = 1$. Suponha ainda que $f_i(\mathbf{x})$ denota a densidade de \mathbf{X} para uma observação que vem do i -ésimo grupo. Portanto, pelo teorema de Bayes temos que

$$\delta_j(\mathbf{x}) = \frac{\pi_j f_j(\mathbf{x})}{\sum_{i=1}^g \pi_i f_i(\mathbf{x})}, \quad j = 1, 2, \dots, g. \quad (3.1)$$

Conseqüentemente, o objeto é alocado para o grupo que fornecer o maior valor de $\delta_j(\mathbf{x})$. Frequentemente, o valor de π_i pode ser estimado pela fração amostral de cada grupo (JAMES *et al.*, 2013, p. 139)

Pode-se ver que, aparentemente, é um método bastante simples, mas existem dois problemas envolvidos. O primeiro, é escolher uma distribuição para os dados. O segundo, estimar os parâmetros dessa distribuição.

3.2.1 Máxima Verossimilhança

Geralmente quando se supõe uma distribuição para os dados, não conhecemos os parâmetros dessa distribuição, portanto, sendo necessário estimá-los. O método da máxima verossimilhança (MV) foi proposto por Ronald Aylmer Fisher em 1912, no qual o trabalho foi reimpresso em Fisher (1992), e é um dos métodos mais utilizados para estimação de parâmetros na Estatística. Posteriormente, Aldrich (1997) considerou as mudanças de justificativas de Fisher para o método, os conceitos que ele desenvolveu em torno dele (incluindo a verossimilhança, suficiência, eficiência e informação) e as abordagens que ele descartou. Sejam $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ uma amostra aleatória de tamanho n de uma variável aleatória \mathbf{X} , com função de densidade ou de probabilidade $f(\mathbf{x}; \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor de parâmetros com $\boldsymbol{\theta} \in \Theta$ (espaço paramétrico). A função de verossimilhança da amostra observada é dada por

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta}). \quad (3.2)$$

Para uma dada amostra \mathbf{x} , seja $\hat{\boldsymbol{\theta}}(\mathbf{x})$ o valor do vetor de parâmetros que maximiza (3.2) como função de $\boldsymbol{\theta}$. Temos que o estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ com base em \mathbf{X} é o vetor $\hat{\boldsymbol{\theta}}(\mathbf{X}) \in \Theta$. Habitualmente, para encontrar o EMV são utilizadas as técnicas de derivação.

Assim, utilizando o fato de que o logaritmo natural é uma função monotonicamente crescente, é mais fácil maximizar o logaritmo de (3.2), pois derivar uma soma é mais conveniente do que um produto. O logaritmo natural da verossimilhança é

$$\begin{aligned}\ell(\boldsymbol{\theta}; \mathbf{x}) &= \ln(L(\boldsymbol{\theta}; \mathbf{x})) \\ &= \sum_{i=1}^n \ln(f(\mathbf{x}_i; \boldsymbol{\theta})).\end{aligned}$$

Em alguns exemplos simples, o EMV possui uma forma explícita. Em situações mais complicadas, o que comumente acontece, o EMV é encontrado através de métodos numéricos. Pode-se dizer que o EMV é o mais usado em problemas de estimação, devido às suas diversas propriedades como: convergência em probabilidade, convergência em distribuição para a normal, é função de estatísticas suficientes, é assintoticamente não-viesado, dentre outras. Para mais detalhes sobre as propriedades e exemplos da MV e do EVM, ver Bolfarine e Sandoval (2010) e Casella e Berger (2002). À vista disso, o método da MV será utilizado para todas as distribuições que forem consideradas no presente trabalho.

3.2.2 Discriminante - Bingham Complexa

A distribuição Bingham Complexa foi proposta por Kent (1994) e é muito utilizada em análise de formas devido a sua simetria rotacional. Considere uma variável aleatória $\mathbf{Z} \in \mathbb{C}S^{k-2}$ (definido na equação (2.4)) tendo distribuição Bingham complexa, diga-se $\mathbf{Z} \sim \mathbb{CB}_{k-2}(\mathbf{A})$, que possui densidade

$$f(\mathbf{z}) = c(\mathbf{A})^{-1} \exp(\mathbf{z}^* \mathbf{A} \mathbf{z}),$$

na qual \mathbf{A} é uma matriz Hermitiana e $c(\mathbf{A})$ é a constante de normalização dada por

$$c(\mathbf{A}) = 2\pi^{k-1} \sum_{i=1}^{k-1} a_i \exp(\tilde{\lambda}_i), \quad a_i^{-1} = \prod_{i \neq j} (\tilde{\lambda}_j - \tilde{\lambda}_i),$$

em que $\tilde{\lambda}_1 < \tilde{\lambda}_2 < \dots < \tilde{\lambda}_{k-1} = 0$ são os autovalores da matriz \mathbf{A} .

Suponha que temos duas populações independentes \mathbf{Z}_1 e \mathbf{Z}_2 tais que, $\mathbf{Z}_1 \sim \mathbb{CB}_{k-2}(\mathbf{A}_1)$ e $\mathbf{Z}_2 \sim \mathbb{CB}_{k-2}(\mathbf{A}_2)$. Note que, quando temos apenas duas populações ($g = 2$), como alternativa, basta calcular a razão entre as densidades das duas populações. Portanto, a função discriminante de Bayes é dada por

$$\delta(\mathbf{z}) = \frac{\pi_1 c(\mathbf{A}_1)^{-1} \exp(\mathbf{z}^* \mathbf{A}_1 \mathbf{z})}{\pi_2 c(\mathbf{A}_2)^{-1} \exp(\mathbf{z}^* \mathbf{A}_2 \mathbf{z})} = \frac{\pi_1 c(\mathbf{A}_2)}{\pi_2 c(\mathbf{A}_1)} \exp\{\mathbf{z}^* (\mathbf{A}_1 - \mathbf{A}_2) \mathbf{z}\}.$$

Dada uma nova observação \mathbf{z}' , se $\delta(\mathbf{z}') > 1$, então \mathbf{z}' é alocado para o grupo 1; se $\delta(\mathbf{z}') < 1$, \mathbf{z}' é alocado para o grupo 2; por fim, caso $\delta(\mathbf{z}') = 1$, o objeto não é alocado para nenhum dos grupos. Esse classificador foi utilizado por Southworth, Mardia e Taylor (2000) em dados sobre vértebras de ratos.

3.2.2.1 Inferência: Bingham Complexa

Em relação à estimação dos parâmetros da distribuição Bingham complexa, considere uma amostra $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ tendo distribuição $\mathbb{C}B_{k-2}(\mathbf{A})$. Assim, deve-se maximizar

$$\begin{aligned} \ell(\mathbf{A}; \mathbf{z}) &= \sum_{i=1}^n \ln(f(\mathbf{z}_i; \mathbf{A})) = \sum_{i=1}^n \mathbf{z}_i^* \mathbf{A} \mathbf{z}_i - n \ln(c(\mathbf{A})) \\ &= \text{traço} \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^* \mathbf{A} \right) - n \ln(c(\mathbf{A})) \\ &= \text{traço}(\mathbf{S}\mathbf{A}) - n \ln(c(\mathbf{A})). \end{aligned}$$

Considere a decomposição espectral (Teorema 1) da matriz $\mathbf{A} = \sum_{i=1}^{k-1} \tilde{\lambda}_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^*$ e $\mathbf{S} = \sum_{i=1}^{k-1} \lambda_i \mathbf{g}_i \mathbf{g}_i^*$. Dessa forma, tem-se que

$$\ell(\mathbf{A}; \mathbf{z}) = \text{traço} \left(\sum_{i=1}^{k-1} \tilde{\lambda}_i \boldsymbol{\gamma}_i \mathbf{S} \boldsymbol{\gamma}_i^* \right) - n \ln(c(\mathbf{A})).$$

O primeiro termo de $\ell(\mathbf{A}; \mathbf{z})$, é maximizado utilizando o resultado do Teorema 2, ou seja, o máximo de $\boldsymbol{\gamma}_i \mathbf{S} \boldsymbol{\gamma}_i^* = \lambda_i$, quando $\boldsymbol{\gamma}_i = \mathbf{g}_i$, portanto, o EMV de $\boldsymbol{\gamma}_i$ é

$$\hat{\boldsymbol{\gamma}}_i = \mathbf{g}_i, \quad i = 1, 2, \dots, k-1.$$

Assim,

$$\ell(\mathbf{A}; \mathbf{z}) = \sum_{i=1}^{k-1} \lambda_i \tilde{\lambda}_i - n \ln(c(\mathbf{A})). \quad (3.3)$$

Derivando (3.3) em relação a cada $\tilde{\lambda}_i$ e igualando a zero, dispomos que

$$\frac{\partial \ell(\mathbf{A}; \mathbf{z})}{\partial \tilde{\lambda}_i} = \lambda_i - n \frac{\partial \ln(c(\mathbf{A}))}{\partial \tilde{\lambda}_i} = 0,$$

o que implica em

$$\frac{\partial \ln(c(\mathbf{A}))}{\partial \tilde{\lambda}_i} = \frac{1}{n} \lambda_i. \quad (3.4)$$

Note que (3.4) é uma equação não-linear e deve ser resolvida através de métodos numéricos. Uma alternativa para esse problema, é considerar alta concentração nos dados, ou seja, assumir que o valor de $\lambda_1 \cong n$ (KENT, 1994), que implica que os autovalores de \mathbf{A} são grandes em valor absoluto. Desse modo, os estimadores de MV de $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{k-2}$ são aproximadamente

$$\hat{\tilde{\lambda}}_i \cong -\frac{n}{\lambda_i}, \quad i = 1, 2, \dots, k-2.$$

3.2.3 Discriminante - Watson Complexa

A distribuição Watson complexa também é definida no espaço das pré-formas $\mathbb{C}S^{k-2}$. Ela é denotada por $\mathbb{C}W_{k-2}(\boldsymbol{\mu}, \kappa)$, e uma variável aleatória \mathbf{Z} é dita ter distribuição Watson se possuir a densidade

$$f(\mathbf{z}) = c(\kappa)^{-1} \exp \{ \kappa |\mathbf{z}^* \boldsymbol{\mu}|^2 \}, \quad (3.5)$$

em que $\boldsymbol{\mu}$ é a forma média, $\kappa \geq 0$ representa o parâmetro de concentração e $c(\kappa)$ é a constante de normalização calculada por (MARDIA; DRYDEN, 1999)

$$c(\kappa) = 2\pi^{k-1} \kappa^{2-k} \left\{ e^\kappa - \sum_{j=0}^{k-3} \frac{\kappa^j}{j!} \right\}. \quad (3.6)$$

Essa distribuição é uma caso particular da Bingham complexa, isso acontece quando há apenas dois autovalores distintos em \mathbf{A} (um único autovalor distinto e todos os outros autovalores iguais), ou seja, fazendo $\mathbf{A} = -\kappa(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$. Algumas propriedades, como: estimação, testes de hipóteses e aproximações da distribuição Watson complexa, foram discutidas em Mardia e Dryden (1999).

Suponha, novamente, que temos duas populações independentes \mathbf{Z}_1 e \mathbf{Z}_2 tais que, $\mathbf{Z}_1 \sim \mathbb{C}W_{k-2}(\boldsymbol{\mu}_1, \kappa_1)$ e $\mathbf{Z}_2 \sim \mathbb{C}W_{k-2}(\boldsymbol{\mu}_2, \kappa_2)$. Sejam π_1 e π_2 as probabilidades *a priori* de um objeto pertencer a uma das duas populações, portanto a função discriminante de Bayes é dada por

$$\delta(\mathbf{z}) = \frac{\pi_1 c(\kappa_2)}{\pi_2 c(\kappa_1)} \exp \{ \kappa_1 |\mathbf{z}^* \boldsymbol{\mu}_1|^2 - \kappa_2 |\mathbf{z}^* \boldsymbol{\mu}_2|^2 \}.$$

Dada uma nova observação \mathbf{z}' , se $\delta(\mathbf{z}') > 1$, então \mathbf{z}' é alocado para o grupo 1; se $\delta(\mathbf{z}') < 1$, \mathbf{z}' é alocado para o grupo 2; por fim, caso $\delta(\mathbf{z}') = 1$, o objeto não é alocado para nenhum dos grupos.

3.2.3.1 Inferência: Watson Complexa

Considere uma amostra $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ tendo distribuição $\mathbb{C}W_{k-2}(\boldsymbol{\mu}, \kappa)$. Assim, deve-se maximizar

$$\begin{aligned} \ell(\boldsymbol{\mu}, \kappa; \mathbf{z}) &= \sum_{i=1}^n \ln(f(\mathbf{z}_i; \boldsymbol{\mu}, \kappa)) = \kappa \sum_{i=1}^n |\mathbf{z}_i^* \boldsymbol{\mu}|^2 - n \ln(c(\kappa)) \\ &= \kappa \sum_{i=1}^n \boldsymbol{\mu}^* \mathbf{z}_i \mathbf{z}_i^* \boldsymbol{\mu} - n \ln(c(\kappa)) = \kappa \boldsymbol{\mu}^* S \boldsymbol{\mu} - n \ln(c(\kappa)). \end{aligned}$$

Pelo Teorema 2, o máximo de $\boldsymbol{\mu}^* S \boldsymbol{\mu}$ é igual a λ_1 , quando $\boldsymbol{\mu} = \mathbf{u}_1$. Que é justamente a forma média utilizando a distância procrustes completa (2.10). Assim, temos que o EMV de $\boldsymbol{\mu}$ é

$$\hat{\boldsymbol{\mu}} = \mathbf{u}_1.$$

Assim, falta maximizar

$$\ell(\kappa; \mathbf{z}) = \kappa \lambda_1 - n \ln(c(\kappa)), \quad (3.7)$$

em relação a κ . Derivando (3.7) em relação a κ , temos

$$\frac{\partial \ell(\kappa; \mathbf{z})}{\partial \kappa} = \lambda_1 - n \frac{\frac{\partial c(\kappa)}{\partial \kappa}}{c(\kappa)} = 0.$$

A solução é dada para o valor de κ tal que

$$\frac{\partial c(\kappa)}{\partial \kappa} = c(\kappa) \frac{\lambda_1}{n}. \quad (3.8)$$

Pode-se ver que temos uma equação não-linear, consequentemente, é necessário utilizar algum método numérico, por exemplo, o método de Newton-Raphson. Perceba que é um problema

simples, pois envolve apenas um parâmetro. Ademais, supondo que $\kappa \rightarrow \infty$, então a constante em (3.6) é aproximadamente

$$c(\kappa) \cong 2\pi^{k-1} \kappa^{2-k} e^\kappa,$$

e

$$\frac{\partial c(\kappa)}{\partial \kappa} \cong 2\pi^{k-1} \kappa^{2-k} e^\kappa \left(1 - \frac{k-2}{\kappa}\right).$$

Assim sendo, (3.8) pode ser aproximada por

$$2\pi^{k-1} \kappa^{2-k} e^\kappa \left(1 - \frac{k-2}{\kappa}\right) = 2\pi^{k-1} \kappa^{2-k} e^\kappa \frac{\lambda_1}{n},$$

que implica em

$$\left(1 - \frac{k-2}{\kappa}\right) = \frac{\lambda_1}{n} \rightarrow \kappa = \frac{n(k-2)}{n-\lambda_1}.$$

Portanto, o EMV de κ é

$$\hat{\kappa} \cong \frac{n(k-2)}{n-\lambda_1}.$$

O Algoritmo 1 descreve os passos para aplicação do classificador de Bayes.

Algoritmo 1: Classificador de Bayes.

Entrada: Probabilidades $\pi_1, \pi_2, \dots, \pi_g$, amostras $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in_i}$ de cada população e escolha de $f_i(\cdot)$, com $i = 1, 2, \dots, g$.

Saída: Classificador de Bayes.

1: Estimar os parâmetros via máxima verossimilhança para cada $f_i(\cdot)$ com a utilização da respectiva amostra;

2: Calcular os valores do discriminante de Bayes para uma nova observação \mathbf{x}'

$$\delta_j(\mathbf{x}') = \frac{\pi_j f_j(\mathbf{x}')}{\sum_{i=1}^g \pi_i f_i(\mathbf{x}')}, \quad j = 1, 2, \dots, g;$$

3: Alocar \mathbf{x}' ao grupo que possuir o maior valor de δ_j .

3.3 SUPPORT VECTOR MACHINE

O classificador Máquina de Suporte Vetorial (*Support Vector Machine* - SVM) foi proposto por Boser, Guyon e Vapnik (1992) e desde então, vem sendo utilizado em diversas áreas

em problemas de classificação, regressão, agrupamento, etc. A SVM é um classificador binário com aprendizado supervisionado, ou seja, é necessário fornecer exemplos dos dois grupos que definem o que deverá ser buscado pelo algoritmo. A ideia da SVM é encontrar um hiperplano que separa os dois grupos, satisfazendo o critério de maior margem. Ou seja, podemos calcular a distância (perpendicular) de cada observação de treinamento para um determinado hiperplano de separação; a menor distância é a distância mínima das observações até o hiperplano e é conhecida como a margem (JAMES *et al.*, 2013, p. 341).

3.3.1 Dados Linearmente Separáveis

Um conjunto de dados é dito ser linearmente separável se é possível separar completamente os padrões dos grupos diferentes por pelo menos um hiperplano. Por definição, a equação do hiperplano é dada por

$$h(\mathbf{x}) = \mathbf{w}_h^\top \mathbf{x} + b_h = 0, \quad (3.9)$$

em que $\mathbf{w}_h \in \mathbb{R}^p$ é o vetor de coeficientes, $\mathbf{x} \in \mathbb{R}^p$ é o vetor de variáveis e b_h representa um escalar. Note que esse hiperplano divide o espaço em duas regiões, $h(\mathbf{x}) > 0$ e $h(\mathbf{x}) < 0$. Como ilustração observe a Figura 6, em que os dois grupos estão sendo separados por um hiperplano, para mais detalhes ver Lorena e Carvalho (2007). Perceba que na Figura 6, está representado apenas um possível hiperplano, pois poderia-se ter infinitos hiperplanos apenas multiplicando \mathbf{w}_h por uma constante, ou ainda, somando-a em b_h como mostra na Figura 7.

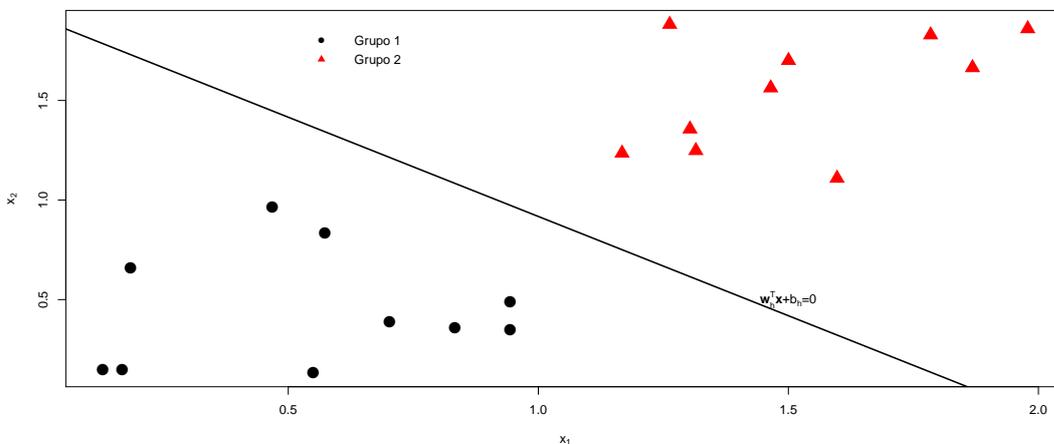


Figura 6 – Dados linearmente separáveis com um hiperplano separador.

Portanto, o critério proposto por Boser, Guyon e Vapnik (1992) consiste em estimar \mathbf{w}_h e b_h que maximizam a distância entre as margens. Com isso, maximizando a distância entre

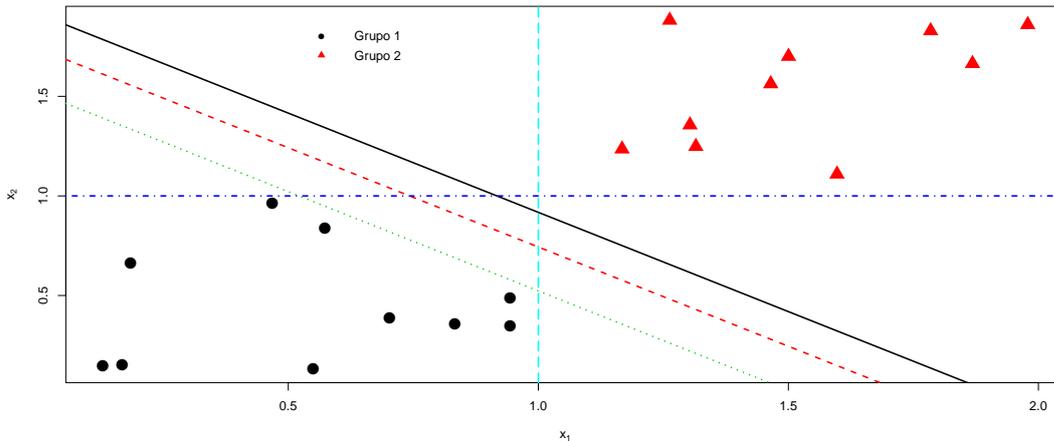


Figura 7 – Dados linearmente separáveis com 5 hiperplanos separadores.

os grupos e, por conseguinte, aumentando o poder de generalização do classificador. Na Figura 8 pode-se ver o hiperplano que separa os dois grupos utilizando a abordagem de maior margem. As retas pontilhadas são as margens que possuem a maior distância entre os grupos e os pontos que estão sobre as margens são chamados de vetores de suporte (VS). Os VS são utilizados no cálculo para a obtenção do hiperplano.

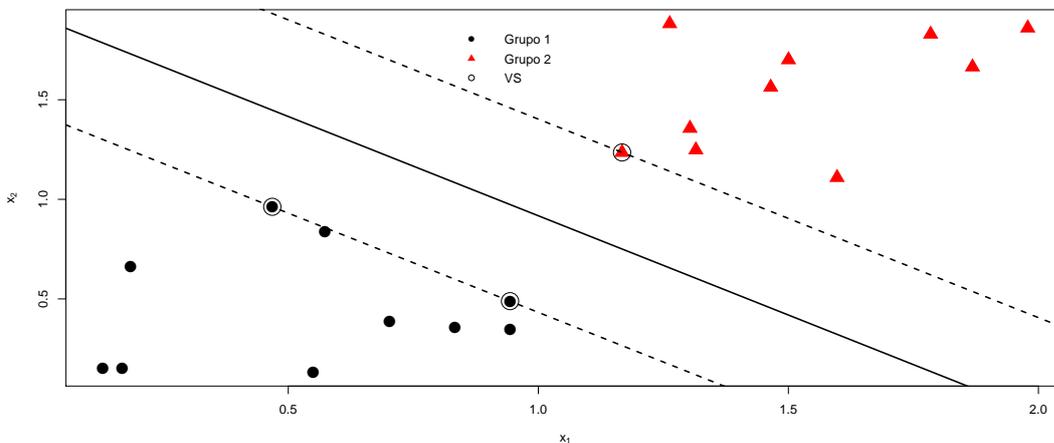


Figura 8 – Dados linearmente separáveis com o hiperplano de margem máxima.

3.3.2 SVM Lineares

Seja um conjunto de amostras de treinamento com dois grupos (1 e 2), representado pelo conjunto de vetores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, em que cada vetor possui o rótulo, $y_i = +1$ se \mathbf{x}_i pertencer ao grupo 1, e $y_i = -1$ se \mathbf{x}_i pertencer ao grupo 2, ou ainda, podemos rotular as amostras da

seguinte maneira

$$y_i = \begin{cases} +1 & \text{se } \mathbf{w}_h^\top \mathbf{x}_i + b_h > 0, \\ -1 & \text{se } \mathbf{w}_h^\top \mathbf{x}_i + b_h < 0. \end{cases} \quad (3.10)$$

Repare que no caso dos dados serem linearmente separados, não existe o caso em que $h(\mathbf{x}) = 0$, logo as condições de rotulação em (3.10) podem ser alteradas para

$$y_i = \begin{cases} +1 & \text{se } \mathbf{w}_h^\top \mathbf{x}_i + b_h > a, \\ -1 & \text{se } \mathbf{w}_h^\top \mathbf{x}_i + b_h < -a, \end{cases} \quad (3.11)$$

sendo $a > 0$ uma constante. Veja que segundo o sistema (3.11) não há pontos entre os hiperplanos $\mathbf{w}_h^\top \mathbf{x} + b \pm a$. Com isso, as inequações em (3.11) podem ser reescritas dividindo ambos os lados por a e reajustando \mathbf{w}_h e b_h temos

$$y_i = \begin{cases} +1 & \text{se } \mathbf{w}^\top \mathbf{x}_i + b > 1, \\ -1 & \text{se } \mathbf{w}^\top \mathbf{x}_i + b < -1. \end{cases} \quad (3.12)$$

Essas inequações (3.12) são utilizadas como restrições na função objetivo para encontrar o hiperplano ótimo, pois elas garantem que não haja dados de treinamento entre as margens. Ademais, essas restrições podem ser simplificadas da seguinte forma

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad i = 1, 2, \dots, n. \quad (3.13)$$

Agora é necessário calcular a distância entre as margens, pois o objetivo é maximizar a margem sem que nenhum ponto fique entre elas, por isso é necessário utilizar as inequações em (3.13). A distância entre as margens será calculada utilizando a distância euclidiana. Dessa forma, considere os vetores \mathbf{x}_1 e \mathbf{x}_2 . Assim a distância euclidiana é definida por

$$d_E = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^\top (\mathbf{x}_1 - \mathbf{x}_2)} = \|\mathbf{x}_1 - \mathbf{x}_2\|.$$

Pode-se mostrar que a distância entre qualquer ponto e o hiperplano (3.9) é dada por

$$d(\mathbf{x}_i, h(\mathbf{x})) = \frac{|h(\mathbf{x}_i)|}{\|\mathbf{w}\|} = \frac{|\mathbf{w}^\top \mathbf{x}_i + b|}{\|\mathbf{w}\|},$$

ver Carvalho (2016). Portanto, a distância entre as margens é

$$d(h(\mathbf{x}) = 1, h(\mathbf{x}) = -1) = \frac{2}{\|\mathbf{w}\|}. \quad (3.14)$$

Desta forma, para encontrar os coeficientes basta maximizar (3.14) em relação a \mathbf{w} sujeito as restrições em (3.13). Veja que maximizar (3.14) é equivalente a minimizar $\|\mathbf{w}\|$, ou ainda, equivalentemente, pode-se minimizar $\|\mathbf{w}\|^2$, pois nesse caso a função quadrática é monotonicamente crescente, facilitando o processo de minimização e preservando o mínimo de $\|\mathbf{w}\|$. Portanto, o hiperplano ótimo (de maior margem) é definido para os valores de \mathbf{w} e b que satisfazem (3.13), para os quais $\|\mathbf{w}\|^2$ é mínima. Dessa maneira, o problema pode ser resumido em

$$\min_{\mathbf{w}, b} \frac{\|\mathbf{w}\|^2}{2}, \quad (3.15)$$

sob as restrições $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$ para $i = 1, 2, \dots, n$.

Para resolver o problema de otimização da equação (3.15), pode-se utilizar os multiplicadores de Lagrange. O método permite encontrar extremos (máximo ou mínimo) de uma função de uma ou mais variáveis suscetíveis a uma ou mais restrições. Considere a situação em que

$$\begin{aligned} \min \quad & g_1(\mathbf{x}) \\ \text{sujeito a} \quad & g_2(\mathbf{x}) = c. \end{aligned}$$

O método consiste em introduzir uma variável nova ($\alpha \geq 0$), chamada de multiplicador de Lagrange. A partir disso, estuda-se a função de Lagrange, assim definida

$$L = g_1(\mathbf{x}) - \alpha(g_2(\mathbf{x}) - c).$$

Nessa função, α pode ser subtraído ou somado (sem perda de generalidade). Se existirem mais restrições, basta adicionar mais multiplicadores. Para mais detalhes sobre a teoria envolvida ver Santos (2002). Utilizando os multiplicadores de Lagrange obtém-se a função primal

$$L_{\text{primal}} = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1], \quad (3.16)$$

no qual cada $\alpha_i \geq 0$ com $i = 1, 2, \dots, n$. Essa função deve ser minimizada em relação a \mathbf{w} e b e maximizada em relação aos α_i 's, assim, derivando (3.16) em relação a \mathbf{w} e b temos

$$\frac{\partial L_{\text{primal}}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad (3.17)$$

$$\frac{\partial L_{\text{primal}}}{\partial b} = - \sum_{i=1}^n \alpha_i y_i. \quad (3.18)$$

Igualando as equações (3.17) e (3.18) a 0 e desenvolvendo os cálculos, dispomos que

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad (3.19)$$

$$\sum_{i=1}^n \alpha_i y_i = 0. \quad (3.20)$$

Substituindo (3.19) e (3.20) em (3.16), obtém-se o seguinte problema de otimização

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j, \quad (3.21)$$

$$\text{sob as restrições } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases}.$$

Essa formulação da equação (3.21) é denominada de forma dual e deverá ser maximizada em relação aos α_i 's. Essa estratégia tornou-se padrão para estimar os parâmetros da SVM, pois fornece alguns benefícios, como a redução da alta dimensionalidade dos dados. Considerando $\hat{\alpha}_i$, $\hat{\mathbf{w}}$ e \hat{b} como a solução para o problema, esses valores só serão ótimos se satisfizerem o teorema de Karush-Kuhn-Tucker (KKT) (SANTOS, 2002). As condições de KKT aplicadas ao problema são

$$\begin{aligned} \frac{\partial L_{\text{primal}}}{\partial \mathbf{w}} &= 0, \\ \frac{\partial L_{\text{primal}}}{\partial b} &= 0, \\ y_i(\mathbf{w}^\top \mathbf{x}_i + b) &\geq 1 \quad \forall i, \\ \alpha_i &\geq 0 \quad \forall i, \\ \alpha_i[y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1] &= 0 \quad \forall i. \end{aligned} \quad (3.22)$$

É importante ressaltar que os VS são os únicos que possuem $\hat{\alpha}_i > 0$, com isso, os coeficientes do hiperplano são dados por

$$\hat{\mathbf{w}} = \sum_{i \in B} \hat{\alpha}_i y_i \mathbf{x}_i, \quad (3.23)$$

em que $B = \{l; \hat{\alpha}_l > 0, l = 1, \dots, n\}$. Veja que \mathbf{w} é calculado explicitamente pela equação (3.23). Para calcular b é necessário utilizar a equação (3.22), isolando b e depois calculando a média aritmética das imagens dos VS, ficando na forma

$$\hat{b} = \frac{1}{n_{vs}} \sum_{i \in B} \left(y_i - \sum_{j \in B} y_j \hat{\alpha}_j \mathbf{x}_j^\top \mathbf{x}_i \right), \quad (3.24)$$

no qual n_{vs} é o número de VS. A SVM mostrada aqui é geralmente conhecida como SVM de margens rígidas.

3.3.3 SVM Com Erros

Na grande maioria dos casos os dados podem não ser linearmente separáveis e isso pode acontecer por diversos motivos, por exemplo: ruídos, erros de medição e até mesmo a natureza dos dados. Assim, a abordagem descrita anteriormente não se aplica e uma nova formulação precisa ser feita. Observe que podem acontecer 3 situações que não caracterizam o caso de dados linearmente separáveis. Essas situações estão listadas a seguir,

1. Pontos podem ficar entre as margens e serem classificados corretamente;
2. Pontos podem ficar entre as margens e serem classificados erroneamente;
3. Pontos que estejam fora das margens e classificados erroneamente.

Essas situações estão ilustradas na Figura 9, (a), (b) e (c), respectivamente. Evidentemente, poderia-se ter também, o caso da junção dessas situações.

A solução para esse problema foi proposta por Cortes e Vapnik (1995), no qual é acrescentado uma variável folga $\xi_i \geq 0$ nas restrições (3.12), tornando-as

$$y_i = \begin{cases} +1 & \text{se } \mathbf{w}^\top \mathbf{x}_i + b > 1 - \xi_i, \\ -1 & \text{se } \mathbf{w}^\top \mathbf{x}_i + b < -1 + \xi_i. \end{cases}$$

Combinando essas equações, temos

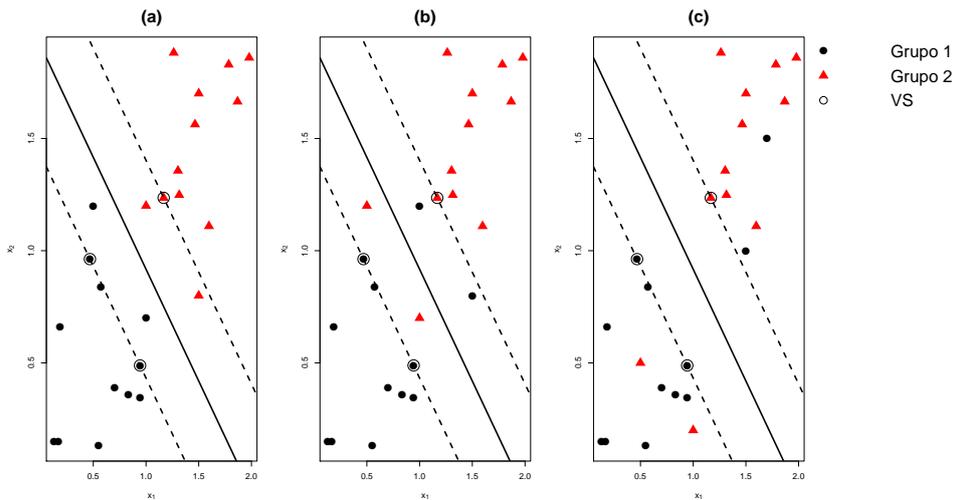


Figura 9 – (a) Encontram-se pontos entre as margens e corretamente classificados; (b) Encontram-se pontos do lado incorreto, mas entre as margens; (c) Encontram-se pontos do lado incorreto e fora das margens.

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \text{ com } \xi_i \geq 0 \text{ e } i = 1, 2, \dots, n.$$

Logicamente, para termos um bom classificador é necessário minimizar os erros, ou melhor, minimizar $\sum_{i=1}^n \xi_i$. Desse modo, o problema para encontrar os coeficientes do hiperplano que possui a maior margem e a menor soma dos erros é dado por

$$\min_{\mathbf{w}, b, \xi_i} \frac{\|\mathbf{w}\|^2}{2} + C \left(\sum_{i=1}^n \xi_i \right), \quad (3.25)$$

sob as restrições $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ para $i = 1, 2, \dots, n$. A constante C simboliza o peso da soma dos erros no processo de minimização e deve ser fornecida pelo usuário. Quanto maior for o valor de C , menor serão as margens do hiperplano. Caso os dados sejam linearmente separáveis, basta escolher $C \rightarrow \infty$ e o resultado será equivalente ao da SVM com margens rígidas.

Para encontrar os coeficientes do hiperplano para o novo problema, pode-se utilizar os multiplicadores de Lagrange do mesmo modo como na abordagem anterior. Assim, a função primal para o novo problema (3.25) é dada por

$$L_{\text{primal}} = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_{i=1}^n \xi_i \right) - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - (1 - \xi_i)] - \sum_{i=1}^n \eta_i \xi_i, \quad (3.26)$$

nos quais α_i e η_i são os multiplicadores de Lagrange. Derivando (3.26) em relação a \mathbf{w} , b e ξ_i e igualando a zero cada uma das derivadas parciais, obtém-se as seguintes equações

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad (3.27)$$

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad (3.28)$$

$$\alpha_i = C - \eta_i. \quad (3.29)$$

Substituindo (3.27), (3.28) e (3.29) em (3.26) consegue-se a função dual dada por

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j, \\ \text{sob as restrições } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C. \end{cases} \end{aligned} \quad (3.30)$$

Note que a forma dual para esse novo problema é similar à formulação para a SVM com margens rígidas, exceto pela restrição de que cada α_i deve ter um valor menor ou igual que a constante C . Ademais, para que a solução desse novo problema seja ótima, as estimativas de $\hat{\alpha}_i$, $\hat{\mathbf{w}}$ e \hat{b} devem atender às condições complementares do teorema de KKT, que para esse caso são

$$\begin{aligned} \frac{\partial L_{\text{primal}}}{\partial \mathbf{w}} &= 0, \\ \frac{\partial L_{\text{primal}}}{\partial b} &= 0, \\ \frac{\partial L_{\text{primal}}}{\partial \xi_i} &= 0 \quad i = 1, 2, \dots, n, \\ y_i(\mathbf{w}^\top \mathbf{x}_i + b) &\geq 1 - \xi_i \quad \forall i, \\ \xi_i &\geq 0 \quad \forall i, \\ \alpha_i &\geq 0 \quad \forall i, \\ \eta_i &\geq 0 \quad \forall i, \\ \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - (1 - \xi_i)] &= 0 \quad \forall i, \\ \eta_i \xi_i &= 0 \quad \forall i. \end{aligned}$$

Por fim, a solução para o problema é

$$\hat{\mathbf{w}} = \sum_{i \in B} \hat{\alpha}_i y_i \mathbf{x}_i,$$

em que $B = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\}$. O valor de \hat{b} é o mesmo como em (3.24). Devido a flexibilidade, essa abordagem proposta por Cortes e Vapnik (1995) ficou conhecida como SVM com margens suaves.

Após encontrar as estimativas do hiperplano, pode-se criar uma função de decisão $De(\mathbf{x}) = \text{sign}(h(\mathbf{x}))$ para a classificação, em que dada uma nova observação \mathbf{x} ,

$$De(\mathbf{x}) = \text{sign}\left(\hat{\mathbf{w}}^\top \mathbf{x} + \hat{b}\right), \quad (3.31)$$

no qual $\text{sign}(\cdot)$ retorna o sinal de um número real, ou seja, o sinal do número é 1, 0 ou -1, se o número é positivo, zero ou negativo, respectivamente.

O Algoritmo 2 resume os passos para o treinamento da SVM com margens suaves.

Algoritmo 2: SVM com margens suaves.

Entrada: Conjunto de treinamento $T = \{\mathbf{x}_i, y_i\}_{i=1}^n$ e o valor de C .

Saída: Classificador SVM.

1: Seja $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n)$ a solução do sistema

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j,$$

$$\text{sob as restrições} \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C. \end{cases}$$

2: O par $(\hat{\mathbf{w}}, \hat{b})$ apresentado a seguir define o hiperplano ótimo

$$\hat{\mathbf{w}} = \sum_{i \in B} \hat{\alpha}_i y_i \mathbf{x}_i, \quad B = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\},$$

$$\hat{b} = \frac{1}{n_{vs}} \sum_{i \in B} \left(y_i - \sum_{j \in B} y_j \hat{\alpha}_j \mathbf{x}_j^\top \mathbf{x}_i \right).$$

3: Dada uma nova observação \mathbf{x}' , a classificação é feita através de

$$De(\mathbf{x}) = \text{sign}\left(\hat{\mathbf{w}}^\top \mathbf{x} + \hat{b}\right).$$

3.3.4 SVM Não-Linear

Muitas vezes na prática os dados não são linearmente separáveis e a abordagem proposta anteriormente pode gerar resultados insatisfatórios, apesar da sua versatilidade. Como exemplo, veja a Figura 10 no qual foi utilizada a SVM com $C = 1000$ em dados simulados. A taxa de acerto para esses dados foi de 65,52%, no qual foram utilizados os próprios dados

de treinamento para a classificação. Repare que seria mais interessante utilizar outra função para separar esses dados, em vez do hiperplano definido em (3.9). Poderia-se utilizar uma transformação para que os dados se tornem linearmente separáveis ou diminua o número de erros. Com o propósito de fornecer essas alternativas, será apresentado um conjunto de transformações chamadas de kernels, que poderá melhorar o desempenho da SVM quando os dados não são linearmente separáveis.

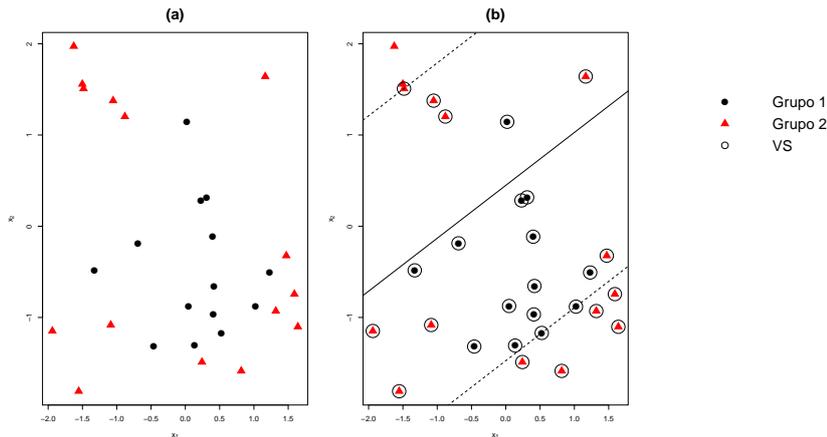


Figura 10 – (a) visualização dos dados; (b) separação dos grupos por um hiperplano.

3.3.4.1 Kernels

A SVM também lida com problemas não lineares, isto é, é possível realizar um mapeamento no espaço original dos dados, chamado de espaço de entradas, para um espaço no qual a dimensão pode ser maior, chamado de espaço de características. Esse mapeamento pode ser representado com a seguinte função

$$\Phi: \mathcal{X} \longrightarrow \mathfrak{S},$$

em que \mathcal{X} denota o espaço de entradas e \mathfrak{S} o espaço de características. A escolha apropriada de Φ pode fazer com que os dados de treinamento sejam linearmente separáveis. Como ilustração, foram utilizados os dados da Figura 10 que estão no \mathbb{R}^2 através de $\mathbf{x} = (x_1, x_2)^\top$, no qual é possível realizar uma transformação para o \mathbb{R}^3 com a função

$$\phi(\mathbf{x}) = (x_1^2, x_1x_2, x_2^2)^\top.$$

Com isso, a equação do hiperplano (3.9) é modificada para

$$\begin{aligned} h(\mathbf{x}) &= \mathbf{w}_h^\top \phi(\mathbf{x}) + b_h \\ &= w_1x_1^2 + w_2x_1x_2 + w_3x_2^2 + b_h. \end{aligned}$$

Nesse espaço de características, pode-se utilizar a SVM com margens suaves e o resultado irá melhorar. Observe na Figura 11 que os dados agora são linearmente separáveis. Foi utilizada a SVM com $C = 100$ resultando em uma taxa de acerto de 100,0% (que antes foi de 65,52%).

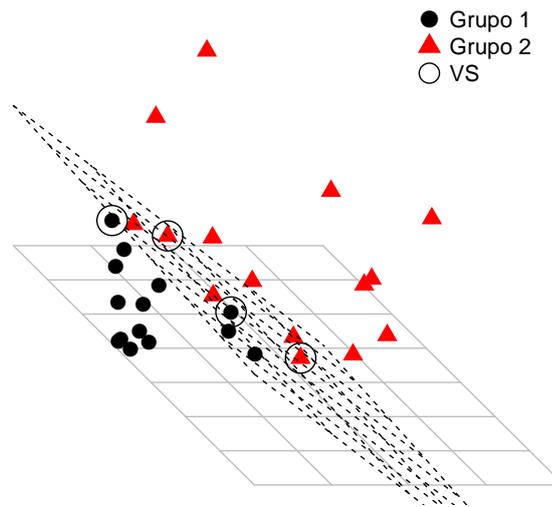


Figura 11 – Dados no espaço de características com o hiperplano separador.

A ideia dessa metodologia é aplicar uma função não linear $\phi(\cdot)$ nas variáveis que estão no espaço de entradas, transformando-as para um espaço de dimensão maior, após isso, utilizar a SVM. Desse modo, um kernel K é definido como uma função que recebe dois argumentos \mathbf{x} e \mathbf{x}' no espaço de entradas e computa o produto $\phi(\mathbf{x})^\top \phi(\mathbf{x}')$ para o espaço de características. Assim, um kernel é representado por

$$K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}').$$

Uma consequência importante dessa representação é que a dimensão do espaço de característica não afeta a complexidade computacional (CRISTIANINI; SHAW-TAYLOR, 2000, p. 30).

Primeiramente, veja que a ideia de um kernel é generalizar o produto interno no espaço de entradas. Claramente, o produto interno fornece um exemplo de kernel, basta fornecer o mapeamento com a função identidade

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}' = \langle \mathbf{x}, \mathbf{x}' \rangle,$$

em que $\langle \cdot, \cdot \rangle$ representa o produto interno entre dois vetores. O objetivo é introduzir a não-linearidade no mapeamento do espaço de entradas para o espaço de características. Portanto, vejamos um simples exemplo, mas ilustrativo, de um tal mapeamento não-linear obtido considerando a seguinte relação

$$\begin{aligned} \langle \mathbf{x}, \mathbf{x}' \rangle^2 &= \left(\sum_{i=1}^p x_i x'_i \right)^2 = \left(\sum_{i=1}^p x_i x'_i \right) \left(\sum_{i=1}^p x_i x'_i \right) \\ &= \sum_{i=1}^p \sum_{j=1}^p x_i x_j x'_i x'_j = \sum_{i,j=(1,1)}^{(p,p)} (x_i x_j) (x'_i x'_j), \end{aligned}$$

que é equivalente a um produto interno entre os vetores

$$\phi(\mathbf{x}) = (x_i x_j)_{i,j=(1,1)}^{(p,p)} \text{ e } \phi(\mathbf{x}') = (x'_i x'_j)_{i,j=(1,1)}^{(p,p)}.$$

Entretanto, nem toda função pode ser considerada um kernel. Para que uma função seja um kernel, é necessário que essa função atenda as condições do Teorema 3. As condições para o Teorema de Mercer são, portanto, equivalentes a exigir que, para qualquer subconjunto finito de \mathcal{X} , a matriz \mathbf{K} correspondente, seja positiva semidefinida, ou seja, seus autovalores são não-negativos (CRISTIANINI; SHAWE-TAYLOR, 2000, p. 33).

Teorema 3 (Teorema de Mercer). Seja \mathcal{X} um espaço de entrada finito com $K(\mathbf{x}, \mathbf{x}')$ sendo uma função simétrica em \mathcal{X} . Então, $K(\mathbf{x}, \mathbf{x}')$ é um kernel se e somente se

$$\mathbf{a}^\top \mathbf{K} \mathbf{a} \geq 0,$$

no qual $\mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^n$ é uma matriz simétrica e $\mathbf{a} \in \mathbb{R}^n$ ($\forall \mathbf{a} \neq 0$).

Na prática, pode-se definir uma função kernel com o espaço de característica implícito. Ou seja, definir uma função kernel para um espaço de entrada é frequentemente mais natural do que criar um espaço de características complicado. Uma maneira para construir novos kernels, é construí-los a partir de kernels mais simples. Isso pode ser feito usando as seguintes propriedades: dado os kernels $K_1(\cdot, \cdot)$ e $K_2(\cdot, \cdot)$, valem-se as seguintes propriedades

1. $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$;
2. $K(\mathbf{x}, \mathbf{x}') = cK_1(\mathbf{x}, \mathbf{x}')$, com $c > 0$;
3. $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}')K_2(\mathbf{x}, \mathbf{x}')$;

4. $K(\mathbf{x}, \mathbf{x}') = g(\mathbf{x})g(\mathbf{x}')$, com $g(\cdot)$ sendo uma função real;
5. $K(\mathbf{x}, \mathbf{x}') = K_3(\phi(\mathbf{x}), \phi(\mathbf{x}'))$, em que $K_3(\cdot, \cdot)$ é um kernel;
6. $K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{A} \mathbf{x}'$, com \mathbf{A} sendo uma matriz semidefinida.

Para mais detalhes ver Cristianini e Shawe-Taylor (2000, p. 42) e Bishop (2006, p. 296). Veja na Tabela 1 alguns exemplos de kernels. Nesse estudo, foram considerados os kernels mais usados na literatura, os kernels: linear, polinomial e Gaussiano.

Tabela 1 – Exemplos de algumas funções do tipo kernel.

Nome do Kernel	$K(\mathbf{x}, \mathbf{x}')$	Parâmetros
Linear	$\mathbf{x}^\top \mathbf{x}'$	-
Polinomial	$[a_1(\mathbf{x}^\top \mathbf{x}') + a_2]^{a_3}$	a_1, a_2 e a_3
Gaussiano ou Radial	$\exp\{-\sigma_r^2 \ \mathbf{x} - \mathbf{x}'\ ^2\}$	σ_r^2
Sigmoide	$\tanh[b_1(\mathbf{x}^\top \mathbf{x}') + b_2]$	b_1 e b_2
Exponencial	$\exp\left\{-\frac{1}{2\sigma_e^2} \ \mathbf{x} - \mathbf{x}'\ \right\}$	σ_e^2

O kernel pode ser inserido diretamente no problema de otimização da SVM para encontrar o hiperplano, ficando na forma

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (3.32)$$

sob as restrições em (3.30). A função de classificação (3.31) é modificada para

$$De(\mathbf{x}) = \text{sign} \left(\sum_{i \in B} y_i \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}) + \hat{b} \right).$$

A SVM também é utilizada quando se tem g grupos ($g > 2$), com o uso dos métodos:

- Um-contra-todos

São construídos g classificadores utilizando a SVM, em que as amostras de treinamento do i -ésimo grupo possuem o rótulo $+1$ e todas as outras amostras com o rótulo -1 . A classificação de uma nova observação \mathbf{x}' é feita pelo classificador binário que produzir o maior valor de $\mathbf{w}_i^\top \mathbf{x}' + b_i$ ($i = 1, 2, \dots, g$).

- Um-contra-um

Nesse método é calculada a SVM para todos os pares possíveis, produzindo $g(g-1)/2$ classificadores binários individuais. A classificação é feita para a classe que obtiver o maior número de votos. O número de classificadores criados pelo um-contra-um é muito maior do que a da abordagem de um um-contra-todos. No entanto, o problema de otimização é menor, o que torna possível um treinamento rápido.

- SVM múltipla

Weston e Watkins (1998) propuseram uma extensão natural, que é distinguir todos os grupos em um único processo de otimização. Existem várias outras extensões, como o *Directed Acyclic Graph SVM* (PLATT; CRISTIANINI; SHAWE-TAYLOR, 2000), *Simplified Multi-class SVM* (SimMSVM) (HSU; LIN, 2002) e uma nova abordagem utilizando o método um-contra-todos (CRAMMER; SINGER, 2001).

Uma comparação entre esses métodos para o problema de múltiplos grupos foi feita por Hsu e Lin (2002), concluindo que os métodos SimMSVM e um-contra-um produziram a maioria dos melhores resultados. O Algoritmo 3 resume os passos para o treinamento da SVM com a utilização de um kernel.

Algoritmo 3: SVM com a utilização de um kernel.

Entrada: Conjunto de treinamento $\phi(T) = \{\phi(\mathbf{x}_i), y_i\}_{i=1}^n$ e o valor de C .

Saída: Classificador SVM.

1: Seja $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n)$ a solução do sistema

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j),$$

$$\text{sob as restrições } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C. \end{cases}$$

2: O par $(\hat{\mathbf{w}}, \hat{b})$ apresentado a seguir define o hiperplano ótimo

$$\hat{\mathbf{w}} = \sum_{i \in B} \hat{\alpha}_i y_i \phi(\mathbf{x}_i), \quad B = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\},$$

$$\hat{b} = \frac{1}{n_{vs}} \sum_{i \in B} \left(y_i - \sum_{j \in B} y_j \hat{\alpha}_j K(\mathbf{x}_i, \mathbf{x}_j) \right).$$

3: Dada uma nova observação \mathbf{x}' , a classificação é feita utilizando

$$De(\mathbf{x}) = \text{sign} \left(\sum_{i \in B} y_i \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}) + \hat{b} \right).$$

3.4 SVM - NÚMEROS COMPLEXOS

Existe uma extensão da SVM para regressão e classificação no espaço dos números complexos, que foi proposta por Bouboulis, Theodoridou e Theodoridis (2013). A princípio, essa extensão foi proposta para a classificação de quatro grupos, na qual os autores intitularam de *Quaternary Classification*, não obstante, facilmente adaptada para dois grupos. A seguir são apresentadas algumas noções sobre derivadas no espaço complexo, pois serão de extrema importância para o entendimento dessa nova abordagem.

3.4.1 Diferenciação Complexa

Analogamente como funções reais, a derivada pode ser definida para funções complexas de variáveis complexas. Seja $g(z) \in \mathbb{C}$, assim

$$g'(z_0) = \lim_{z \rightarrow z_0} \frac{g(z) - g(z_0)}{z - z_0}.$$

O limite acima precisa existir para as infinitas séries $\{z_n\}$ que se aproximam de z_0 , isto é, $\lim_{n \rightarrow \infty} z_n = z_0$. Se $g'(z)$ existe em uma região $R \subset \mathbb{C}$, a função $g(z)$ é chamada de analítica, holomórfica ou regular em R (FISCHER, 2002, p. 406). A seguir, serão mostradas as relações entre derivadas reais e complexas.

Uma função complexa pode ser decomposta em duas funções reais, cada uma dependendo de duas variáveis reais x e y , que são as partes real e imaginária de z , respectivamente. Portanto,

$$g(z) = g(x + iy) = u(x, y) + iv(x, y), \text{ com } z = x + iy.$$

Pode ser mostrado que, para $g(z)$ ser holomórfica, as funções $u(x, y)$ e $v(x, y)$ precisam satisfazer as equações de Cauchy-Riemann (FISCHER, 2002, p. 406). Essas equações são dadas por

$$\begin{aligned} \frac{\partial u(x, y)}{\partial x} &= \frac{\partial v(x, y)}{\partial y}, \\ \frac{\partial v(x, y)}{\partial x} &= -\frac{\partial u(x, y)}{\partial y}. \end{aligned}$$

A derivada complexa de uma função holomórfica pode ser expressada pelas derivadas parciais das funções reais $u(x, y)$ e $v(x, y)$ (FISCHER, 2002, p. 406), assim sendo

$$\frac{\partial g(z)}{\partial z} = \frac{\partial u(x,y)}{\partial x} + i \frac{\partial v(x,y)}{\partial x}.$$

Um problema comum é a otimização de funções reais dependendo de parâmetros complexos. Porém, funções de custo complexo não são de interesse, pois no campo de números complexos não há ordenação e, portanto, minimização ou maximização não faz sentido (FISCHER, 2002, p. 407).

3.4.1.1 Cálculo de Wirtinger

Como alternativa para a derivada complexa, o matemático austríaco Wilhelm Wirtinger (1865-1945) estabeleceu as derivadas parciais de $g(z)$, com respeito a z e z^* , do seguinte modo (FISCHER, 2002, p. 408)

$$\frac{\partial g(z)}{\partial z} \triangleq \frac{1}{2} \left(\frac{\partial g(z)}{\partial x} - i \frac{\partial g(z)}{\partial y} \right),$$

e

$$\frac{\partial g(z)}{\partial z^*} \triangleq \frac{1}{2} \left(\frac{\partial g(z)}{\partial x} + i \frac{\partial g(z)}{\partial y} \right).$$

Em homenagem a Wilhelm Wirtinger, que estabeleceu esse cálculo diferencial, esses resultados são chamados de cálculo de *Wirtinger*. Esse cálculo foi utilizado por Bouboulis, Theodoridou e Theodoridis (2013) para a obtenção do hiperplano para números complexos.

3.4.2 SVM Complexa

Considere agora um hiperplano complexo definido por

$$h_{\mathbb{C}}(\mathbf{z}) = \mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}) + b,$$

em que $\mathbf{w} \in \mathbb{C}^p$, $\phi_{\mathbb{C}}(\cdot)$ é uma função complexa e $b \in \mathbb{C}$. Seja um conjunto de treinamento $\mathbf{z}_j \in \mathbb{C}^p$ para $j = 1, \dots, n$, dividido em quatro grupos G_{++}, G_{+-}, G_{-+} e G_{--} , em que cada observação do conjunto possui um rótulo $y_j \in \{\pm 1 \pm i\}$. Se $y_j = +1 + i$, então a j -ésima observação pertence ao grupo G_{++} , ou seja, $\mathbf{z}_j \in G_{++}$. Se $y_j = -1 + i$, assim $\mathbf{z}_j \in G_{-+}$ e assim sucessivamente para os outros grupos. Nessa abordagem são considerados dois hiperplanos (em vez de um, como no caso real) e duas variáveis de folga, uma para a parte real ξ_j^r e outra para a parte imaginária ξ_j^i .

Assim sendo, o problema primal para a SVM complexa é formulado por

$$\min_{\mathbf{w}, \mathbf{v}, b, C, \xi_j^r, \xi_j^i} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} \|\mathbf{v}\|^2 + C \sum_{j=1}^n (\xi_j^r + \xi_j^i), \quad (3.33)$$

$$\text{sob as restrições} \begin{cases} y_j^r \text{Re}(\mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}_j) + \mathbf{v}^* \phi_{\mathbb{C}}^*(\mathbf{z}_j) + b) \geq 1 - \xi_j^r, \\ y_j^i \text{Im}(\mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}_j) + \mathbf{v}^* \phi_{\mathbb{C}}^*(\mathbf{z}_j) + b) \geq 1 - \xi_j^i, \\ \xi_j^r, \xi_j^i \geq 0, \text{ para } j = 1, \dots, n, \end{cases}$$

em que $y_j^r = \text{Re}(y_j)$, $y_j^i = \text{Im}(y_j)$, $(\cdot)^*$ representa o conjugado e $\mathbf{v} \in \mathbb{C}^p$ são os coeficientes do segundo hiperplano. Para encontrar a solução para o problema em (3.33), também é utilizado o método dos multiplicadores de Lagrange, porém com o auxílio dos resultados mostrados anteriormente sobre derivadas parciais complexas, ou seja, o cálculo de *Wirtinger*. No final, a solução é dada pelas maximizações

$$\max_{\alpha} \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{j=1}^n \sum_{l=1}^n \alpha_l \alpha_j y_l^r y_j^r K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j), \quad (3.34)$$

$$\text{sob as restrições} \begin{cases} \sum_{j=1}^n \alpha_j y_j^i = 0 \\ 0 \leq \alpha_j \leq C \text{ para } j = 1, \dots, n \end{cases},$$

e

$$\max_{\eta} \sum_{j=1}^n \eta_j - \frac{1}{2} \sum_{j=1}^n \sum_{l=1}^n \eta_l \eta_j y_l^i y_j^i K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j), \quad (3.35)$$

$$\text{sob as restrições} \begin{cases} \sum_{j=1}^n \eta_j y_j^i = 0 \\ 0 \leq \eta_j \leq C \text{ para } j = 1, \dots, n \end{cases},$$

em que α_j e η_j são os multiplicadores de Lagrange para as partes real e imaginária, na devida ordem, $K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j) = 2\text{Re}(K_{\mathbb{C}}(\mathbf{z}_l, \mathbf{z}_j))$, com $K_{\mathbb{C}}(\mathbf{z}_l, \mathbf{z}_j) = \phi_{\mathbb{C}}(\mathbf{z}_l)^* \phi_{\mathbb{C}}(\mathbf{z}_j)$, que representa um kernel complexo (BOUBOULIS; THEODORIDOU; THEODORIDIS, 2013). Na Tabela 2 estão alguns exemplos de funções kernels para números complexos.

Note que o problema de otimização em (3.33) transformou-se em dois outros problemas (3.34) e (3.35), como na SVM real (problema dual). As estimativas para os coeficientes são

$$\begin{aligned} \hat{\mathbf{w}} &= \sum_{j \in B_1} (\hat{\alpha}_j y_j^r - i \hat{\eta}_j y_j^i) \phi_{\mathbb{C}}(\mathbf{z}_j), \\ \hat{\mathbf{v}} &= \sum_{j \in B_2} (\hat{\alpha}_j y_j^r - i \hat{\eta}_j y_j^i) \phi_{\mathbb{C}}^*(\mathbf{z}_j). \end{aligned}$$

Tabela 2 – Exemplos de alguns kernels complexos.

Nome do Kernel	$K_{\mathbb{C}}(\mathbf{z}, \mathbf{z}')$	Parâmetros
Linear	$\mathbf{z}^* \mathbf{z}'$	—
Polinomial	$[a_1(\mathbf{z}^* \mathbf{z}') + a_2]^{a_3}$	a_1, a_2 e a_3
Gaussiano ou Radial	$\exp\{-\sigma_G^2 \ \mathbf{z} - \mathbf{z}'\ ^2\}$	σ_G^2

A estimativa do intercepto é $\hat{b} = c^r + ic^i$, no qual c^r e c^i são dados por

$$\hat{c}^r = \frac{1}{n_{vs}^r} \sum_{l \in B_1} (y_l^r - \sum_{j \in B_1} y_j^r \hat{\alpha}_l K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j)),$$

$$\hat{c}^i = \frac{1}{n_{vs}^i} \sum_{l \in B_2} (y_l^i - \sum_{j \in B_2} y_j^i \hat{\eta}_l K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j)),$$

em que $B_1 = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\}$, $B_2 = \{l; 0 < \hat{\eta}_l \leq C, l = 1, \dots, n\}$, n_{vs}^r e n_{vs}^i são os números de vetores de suporte da otimização real e imaginária, respectivamente. Por último, a função de classificação de uma nova observação \mathbf{z}' é dada por

$$De(\mathbf{z}') = \text{sign}_{\mathbb{C}} \left(\sum_{j=1}^n (\hat{\alpha}_j y_j^r + i \hat{\eta}_j y_j^i) K_{\mathbb{C}}^r(\mathbf{z}_j, \mathbf{z}') + \hat{c}^r + i \hat{c}^i \right),$$

no qual $\text{sign}_{\mathbb{C}}(\mathbf{z}) = \text{sign}(\text{Re}(\mathbf{z})) + i \text{sign}(\text{Im}(\mathbf{z}))$. Os cálculos envolvidos e toda a construção da SVM complexa podem ser vistos com mais detalhes em Bouboulis, Theodoridou e Theodoridis (2013).

Segundo os autores, essa abordagem pode ser estendida para dois grupos, que é o nosso interesse, apenas considerando que os rótulos (y_i) assumam apenas os valores +1 ou -1. Dessa forma, é retirado \mathbf{v} de (3.33) e o problema se torna

$$\min_{\mathbf{w}, b, C, \xi_j^r, \xi_j^i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^n (\xi_j^r + \xi_j^i), \quad (3.36)$$

$$\text{sob as restrições} \begin{cases} y_j \text{Re}(\mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}_j) + b) \geq 1 - \xi_j^r, \\ y_j \text{Im}(\mathbf{w}^* \phi_{\mathbb{C}}(\mathbf{z}_j) + b) \geq 1 - \xi_j^i, \\ \xi_j^r, \xi_j^i \geq 0, \text{ para } j = 1, \dots, n. \end{cases}$$

O problema dual para (3.36) é

$$\max_{\boldsymbol{\alpha}} \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{j=1}^n \sum_{l=1}^n \alpha_l \alpha_j y_l y_j K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j), \quad (3.37)$$

$$\text{sob as restrições } \begin{cases} \sum_{j=1}^n \alpha_j y_j = 0 \\ 0 \leq \alpha_j \leq C \text{ para } j = 1, \dots, n \end{cases},$$

e

$$\max_{\boldsymbol{\eta}} \sum_{j=1}^n \eta_j - \frac{1}{2} \sum_{j=1}^n \sum_{l=1}^n \eta_l \eta_j y_l y_j K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j), \quad (3.38)$$

$$\text{sob as restrições } \begin{cases} \sum_{j=1}^n \eta_j y_j = 0 \\ 0 \leq \eta_j \leq C \text{ para } j = 1, \dots, n \end{cases}.$$

Observe que os problemas de otimização em (3.37) e (3.38) são idênticos, consequentemente, as estimativas para cada α_i e η_i são as mesmas. Portanto, sendo necessário fazer apenas uma maximização. Dessa forma, as estimativas dos coeficientes são

$$\hat{\mathbf{w}} = \sum_{j \in B_1} (\hat{\alpha}_j y_j - i \hat{\eta}_j y_j) \phi_{\mathbb{C}}(\mathbf{z}_j),$$

e

$$\hat{c}^r = \frac{1}{n_{vs}^r} \sum_{l \in B_1} (y_l - \sum_{j \in B_1} y_j \hat{\alpha}_l K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j)), \quad (3.39)$$

$$\hat{c}^i = \frac{1}{n_{vs}^i} \sum_{l \in B_2} (y_l - \sum_{j \in B_2} y_j \hat{\eta}_l K_{\mathbb{C}}^r(\mathbf{z}_l, \mathbf{z}_j)). \quad (3.40)$$

O Algoritmo 4 resume os passos para o treinamento da SVM tendo como entrada os números complexos para dois grupos.

Algoritmo 4: SVM para dados no espaço dos números complexos.

Entrada: Conjunto de treinamento $\phi_{\mathbb{C}}(T) = \{\phi_{\mathbb{C}}(\mathbf{z}_i), y_i\}_{i=1}^n$ e o valor de C .

Saída: Classificador SVM complexa.

1: Seja $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n)$ a solução do sistema

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K_{\mathbb{C}}^r(\mathbf{z}_i, \mathbf{z}_j),$$

$$\text{sob as restrições } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C \text{ para } i = 1, \dots, n. \end{cases}$$

2: O par $(\hat{\mathbf{w}}, \hat{b})$ apresentado a seguir define o hiperplano ótimo

$$\hat{\mathbf{w}} = \sum_{j \in B} (\hat{\alpha}_j y_j - i \hat{\alpha}_j y_j) \phi_{\mathbb{C}}(\mathbf{z}_j), \quad B = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\},$$

$$\hat{b} = \hat{c}^r + i \hat{c}^i, \text{ obtidos em (3.39) e (3.40).}$$

3: Dada uma nova observação \mathbf{z}' , basta utilizar a função de decisão dada por

$$De(\mathbf{z}') = \text{sign}_{\mathbb{C}} \left(\sum_{j=1}^n (\hat{\alpha}_j y_j + i \hat{\alpha}_j y_j) K_{\mathbb{C}}^r(\mathbf{z}_j, \mathbf{z}') + \hat{b} \right).$$

3.4.3 SVM com Pesos

O método *boosting* é utilizado para combinar classificadores que não obtiverem um bom desempenho para produzir um classificador melhor. Uma das versões mais populares do método *boosting* é o algoritmo chamado de *AdaBoost.M1* proposto por Freund e Schapire (1997). Basicamente, para cada iteração do método, os pesos atribuídos para cada observação são modificados individualmente e o classificador é reaplicado às observações ponderadas. Os pesos aumentam para aquelas observações que foram classificadas erroneamente, em contra partida, os pesos diminuem para aquelas que foram classificadas corretamente. Assim, as observações que são difíceis de classificar corretamente recebem uma influência cada vez maior, portanto, cada sucessivo classificador é forçado a se concentrar nas observações de treinamento que foram classificadas erroneamente.

Considere uma amostra de treinamento $T = \{\mathbf{x}_i, y_i\}$ com $i = 1, \dots, n$, assim, o algoritmo pode ser descrito pelos seguintes passos

1. Defina o número de interações *boosting* n_b .

2. Inicialize os pesos para cada amostra de treinamento com $w_i = 1/n$ com $i = 1, \dots, n$.
3. Repita para $j = 1$ até n_b :
 - Ajuste o classificador $F_j(\mathbf{x}) \in \{+1, -1\}$ com os pesos w_i no conjunto de treinamento.
 - Calcule:

$$\begin{aligned} erro_j &= \sum_{i=1}^n w_i I_{[y_i \neq F_j(\mathbf{x}_i)]}, \\ l_j &= \log [erro_j / (1 - erro_j)], \end{aligned}$$

em que

$$I_{[y_i \neq F_j(\mathbf{x}_i)]} = \begin{cases} 1 & \text{se } y_i \neq F_j(\mathbf{x}_i), \\ 0 & \text{c.c} \end{cases}.$$

- Atualize os pesos:

$$w_i = w_i \exp \left\{ l_j I_{[y_i \neq F_j(\mathbf{x}_i)]} \right\}, i = 1, \dots, n.$$

Depois, faça $w_i \leftarrow w_i / \sum_i w_i$.

4. Combine os classificadores formando apenas um

$$F(\mathbf{x}) = \text{sign} \left\{ \sum_{j=1}^{n_b} l_j F_j(\mathbf{x}) \right\}.$$

Uma abordagem do *boosting* para a SVM foi proposta por García e Lozano (2007).

Basicamente, os pesos são acrescentados na função objetivo (3.32), da seguinte maneira

$$\min_{\mathbf{w}, b, \xi_i} \frac{\|\mathbf{w}\|^2}{2} + C \left(\sum_{i=1}^n \xi_i w_i \right), \quad (3.41)$$

sob as restrições $y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ para $i = 1, 2, \dots, n$.

O problema dual para (3.41) é

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{sob as restrições} \quad & \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C w_i. \end{cases} \end{aligned}$$

Note que a única diferença dessa forma dual para a convencional, é que a variação de α_i está limitada por $C w_i$. Porém, a SVM é considerada um aprendiz forte (*strong learner*) e o

método *boosting* não parece oferecer vantagens significativas em termos de generalização do erro. Ademais, o desempenho da SVM diminui à medida que o número de iterações aumenta (GARCÍA; LOZANO, 2007).

A proposta é dar um peso maior para as observações que estão mais afastadas da forma média dos objetos. Dessa forma, a SVM se preocupará com os objetos mais distantes e, conseqüentemente, podendo melhorar a classificação. A proposta é calcular os pesos da seguinte maneira,

$$w_i = d(\mathbf{x}_i, \boldsymbol{\mu}), \quad i = 1, 2, \dots, n, \quad (3.42)$$

em que $d(\cdot, \cdot)$ é alguma das distâncias (2.5), (2.6) ou (2.7), normalizadas para que $w_i \in (0, 1)$. Para a distância (2.5), não é necessário realizar essa normalização, pois essa distância já está no intervalo $(0, 1)$. No caso da distância d_P (2.6), basta dividir (3.42) por $\sqrt{2}$. Por fim, na distância geodésica (2.7) é necessário dividir (3.42) por $\pi/2$.

No caso da SVM no espaço dos números complexos, pode-se utilizar as distâncias procrustes completa (2.5), procrustes parcial (2.6) e geodésica (2.7). Assim sendo, a forma dual da SVM complexa com os pesos w_i 's é dada por

$$\begin{aligned} & \max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K_{\mathbb{C}}^r(\mathbf{z}_i, \mathbf{z}_j), \\ \text{sob as restrições} & \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C w_i \text{ para } i = 1, \dots, n. \end{cases} \end{aligned}$$

O Algoritmo 5 descreve os passos para a utilização da SVM ponderada com pesos baseados em distâncias.

Algoritmo 5: SVM para dados no espaço dos números complexos com a utilização de pesos.

Entrada: Conjunto de treinamento $\phi_{\mathbb{C}}(T) = \{\phi_{\mathbb{C}}(\mathbf{x}_i), y_i, w_i\}_{i=1}^n$ e o valor de C .

Saída: Classificador SVM complexa com pesos.

1: Seja $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n)$ a solução do sistema

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \hat{\alpha}_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j K_{\mathbb{C}}^r(\mathbf{z}_i, \mathbf{z}_j),$$

$$\text{sob as restrições } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq w_i C \text{ para } i = 1, \dots, n. \end{cases}$$

2: O par $(\hat{\mathbf{w}}, \hat{b})$ apresentado a seguir define o hiperplano ótimo

$$\hat{\mathbf{w}} = \sum_{j \in B} (\hat{\alpha}_j y_j - i \hat{\alpha}_j y_j) \phi_{\mathbb{C}}(\mathbf{z}_j), \quad B = \{l; 0 < \hat{\alpha}_l \leq C, l = 1, \dots, n\},$$

$$\hat{b} = \hat{c}^r + i \hat{c}^i, \text{ obtidos em (3.39) e (3.40).}$$

3: Dada uma nova observação \mathbf{z}' , basta utilizar a função de decisão dada por

$$De(\mathbf{z}') = \text{sign}_{\mathbb{C}} \left(\sum_{j=1}^n (\hat{\alpha}_j y_j + i \hat{\alpha}_j y_j) K_{\mathbb{C}}^r(\mathbf{z}_j, \mathbf{z}') + \hat{b} \right).$$

3.5 ESTIMAÇÃO DE DENSIDADE

A função de densidade de probabilidade é essencial para a estatística, pois a todo momento é atribuída uma distribuição para os dados. Considere uma variável aleatória X com densidade f . Especificamente, a função f fornece uma natural descrição da variável aleatória X e permite calcular probabilidades através da seguinte relação

$$\Pr(a < X < b) = \int_a^b f(x) dx, \quad \text{para } a < b.$$

Na prática existem duas abordagens para estimar f ,

- **paramétrica:** se supõe que os dados são provenientes de uma família paramétrica conhecida de distribuições e em seguida os parâmetros são estimados.
- **não paramétrica:** diferentemente da primeira abordagem, não se supõe distribuição para os dados. Embora, se pressupõe que os dados possuem uma distribuição que tem densidade f , na qual os dados serão autorizados a falar por si mesmos na determinação da estimativa de f .

O enfoque nesta seção será dado à abordagem não paramétrica.

3.5.1 O Estimador Ingênuo

Considere uma amostra aleatória X_1, \dots, X_n com densidade desconhecida f . Segundo Silverman (1986) o estimador da densidade é dado por

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} w\left(\frac{x - X_i}{h_1}\right), \quad h_1 > 0 \quad (3.43)$$

no qual $w(x) = \begin{cases} \frac{1}{2} \text{ se } |x| < 1 \\ 0 \text{ caso contrário.} \end{cases}$

O parâmetro h_1 é chamado de suavizador e dependendo de sua escolha, se obtém diferentes formas de $\hat{f}(x)$. Aumentando-se h_1 se obtém intervalos maiores, conseqüentemente a representação dos dados será suave, caso contrário, se h_1 for muito pequeno, a representação será muito ruidosa, ver Figura 12.

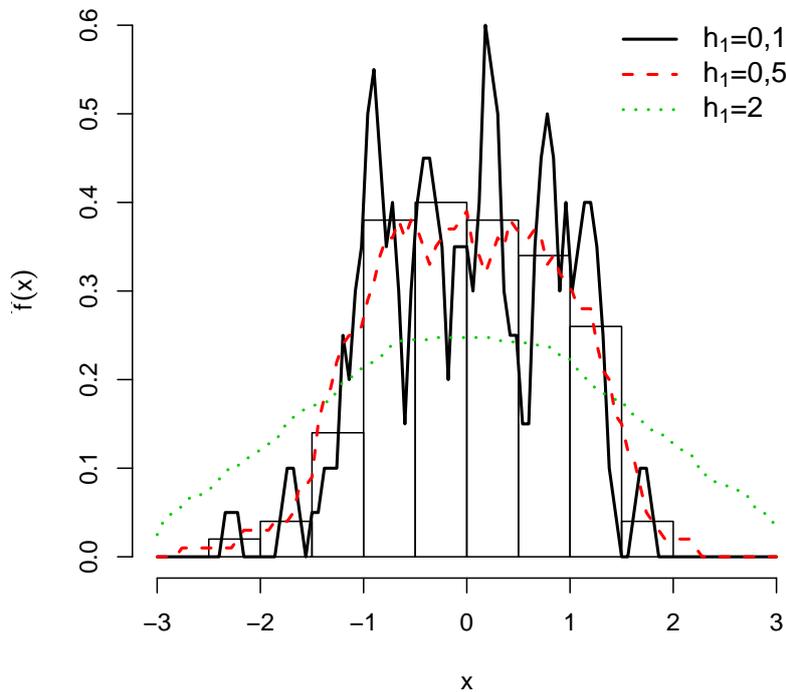


Figura 12 – Histograma de dados simulados sobreposto por estimativas utilizando estimador ingênuo com três valores diferentes de h_1 .

3.5.2 O Estimador Kernel

Segundo Silverman (1986) pode-se generalizar o estimador (3.43), trocando a função peso $w(\cdot)$ por uma função kernel K , tal que

$$\int_{-\infty}^{\infty} K(x) dx = 1.$$

Usualmente, a função $K(x)$ será uma densidade de probabilidade simétrica, como por exemplo, a densidade da distribuição normal. Portanto, o estimador utilizando a função $K(x)$ é

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x - X_i}{h_1}\right).$$

Note que esse kernel é diferente daquele usado na abordagem da SVM, e a principal diferença neste caso, é que a função deve ser uma densidade de probabilidade. Existem vários tipos de kernels univariados, alguns deles estão presentes na Tabela 3 (SILVERMAN, 1986, p. 43).

Tabela 3 – Exemplos de algumas funções kernel para densidades.

Kernel	$K(x)$
Gaussiano	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R}$
Triangular	$1 - x $ para $ x < 1$ e 0 caso contrário
Biweight	$\frac{15}{16}(1 - x^2)^2$ para $ x < 1$, 0 caso contrário
Epanechnikov	$\frac{3}{4} \left(1 - \frac{x^2}{5}\right) / \sqrt{5}$ para $ x < \sqrt{5}$, 0 caso contrário
Retangular	$\frac{1}{2}$ para $ x < 1$, 0 caso contrário

3.5.3 O Estimador Kernel Multivariado

Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ um amostra aleatória multivariada, com cada $\mathbf{X}_i \in \mathbb{R}^p$, extraída de uma mesma população com densidade f . O estimador de f agora é dado por

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1^p} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_1}\right), \quad (3.44)$$

em que $\mathbf{x} = (x_1, \dots, x_p)^\top$ e $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$ com $i = 1, \dots, n$. A função kernel multivariada também precisa ser uma densidade, ou seja,

$$\int_{\mathbb{R}^p} K(\mathbf{x}) dx_1 \dots dx_p = 1.$$

Geralmente a função kernel também será uma densidade unimodal simétrica (SILVERMAN, 1986, p. 76). Um exemplo de kernel multivariado, seria a distribuição normal padrão multivariada

$$K(\mathbf{x}) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{x}\right).$$

Uma outra possibilidade seria o kernel utilizando a distribuição de t-Student com ν graus de liberdade

$$K(\mathbf{x}) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\nu^{p/2}\pi^{p/2}} \left(1 + \frac{1}{\nu}\mathbf{x}^\top \mathbf{x}\right)^{-(\nu+p)/2},$$

nesse caso, o valor de ν deve ser fixado. De modo similar, o estimador de densidade kernel pode ser utilizado para o caso complexo. Os kernels podem ser, por exemplo, as versões complexas da normal multivariada e da t-Student, que possuem as densidades

$$K(\mathbf{x}) = (2\pi)^{-p} \exp\left(-\frac{1}{2}\mathbf{x}^* \mathbf{x}\right),$$

e

$$K(\mathbf{x}) = \frac{\Gamma(\frac{\nu+2p}{2})}{\Gamma(\frac{\nu}{2})\nu^p \pi^p} \left(1 + \frac{1}{\nu}\mathbf{x}^* \mathbf{x}\right)^{-(\nu+2p)/2}.$$

Uma outra alternativa, seria utilizar as distribuições Bingham e Watson complexa. Porém, essas distribuições não possuem uma forma padrão, como na normal e t-Student. Portanto, encontra-se uma certa dificuldade em utilizá-las, visto que, são necessários fornecer muitos parâmetros.

O estimador de densidades será utilizado como no classificador de Bayes (3.1). Em vez de supor uma distribuição f para os dados, será utilizado o estimador não paramétrico \hat{f} (3.44) para o procedimento de classificação. Esse classificador será denominado como Bayes-ED. O Algoritmo 6 possui os passos para esse classificador.

Algoritmo 6: Classificador de Bayes-ED.

Entrada: Probabilidades $\pi_1, \pi_2, \dots, \pi_g$, amostras $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in_i}$ de cada população e escolha de $f_i(\cdot) = K(\cdot)$, com $i = 1, 2, \dots, g$ e o valor de h_1 .

Saída: Classificador de Bayes.

1: Substituir $f_i(\cdot)$ em (3.1) por $\hat{f}_i(\cdot)$ (3.44) para a respectiva amostra.

2: Montar o discriminante de Bayes

$$\delta_j(\mathbf{x}) = \frac{\pi_j \hat{f}_j(\mathbf{x})}{\sum_{i=1}^g \pi_i \hat{f}_i(\mathbf{x})}, \quad j = 1, 2, \dots, g.$$

3.6 HILL-CLIMBING

Em análise numérica, o *Hill-Climbing* (HC) é um algoritmo utilizado em problemas de otimização. Basicamente, o algoritmo inicia em uma solução arbitrária para o problema, então tenta-se encontrar uma solução melhor alterando gradualmente um único elemento da solução. Caso haja uma melhora na solução uma mudança é realizada e, posteriormente, uma nova solução é calculada (FRIEDMAN; RUBIN, 1967). Esse processo é repetido até que nenhuma melhora adicional seja encontrada. Esse algoritmo é muito utilizado em problemas de agrupamento, como no agrupamento de abelhas (FRIEDMAN; RUBIN, 1967), em dados botânicos (RUBIN, 1967) e no agrupamento de plantas, vinhos, e doenças cardíacas (SOUZA *et al.*, 2017).

O algoritmo de HC é descrito em Everitt *et al.* (2011) com os seguintes passos:

1. Construir uma solução inicial para o problema;
2. Alterar cada elemento da solução, produzindo novas soluções;
3. Selecionar a melhor solução;
4. Repetir os passos 2 e 3 até que nenhuma melhora seja encontrada.

Com base no HC, Assis, Souza e Amaral (2018) propuseram um método de agrupamento utilizando estatísticas de testes. A ideia é mover uma observação para cada grupo e se houver uma melhora na estatística de teste, no sentido de rejeitar a hipótese nula, a observação é alocada para o grupo que produziu essa melhora. A proposta é utilizar esse mesmo fundamento para a classificação. Atualmente, as estatísticas de teste mais utilizadas em análise de formas para testar duas médias são: James, Hotelling, Lambda e Goodall.

3.6.1 Testes para Duas Amostras

Considere uma amostra aleatória da população j de tamanho n_j e com forma média $\boldsymbol{\mu}_j$, para $j = 1, 2$. Sejam $\mathbf{z}_{1j}, \dots, \mathbf{z}_{n_jj}$ as pré-formas de cada amostra. Por fim, Considere as seguintes hipóteses

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$\mathbf{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

3.6.1.1 Estatística de Hotelling

Sejam $\mathbf{v}_{1j}, \dots, \mathbf{v}_{n_jj}$ as coordenadas tangentes, obtidas da equação (2.11), de $\mathbf{z}_{1j}, \dots, \mathbf{z}_{n_jj}$, com $j = 1, 2$. Considere ainda que $\mathbf{v}_{1j}, \dots, \mathbf{v}_{n_jj}$ possui distribuição normal multivariada, com vetor de médias $\boldsymbol{\mu}_j$ e matriz de covariâncias comum $\boldsymbol{\Sigma}$. Nos quais, $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2$ são os vetores de médias e $\widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_2$ são as matrizes de covariâncias amostrais em cada grupo que são calculados por

$$\bar{\mathbf{v}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} v_{ij},$$

$$\widehat{\boldsymbol{\Sigma}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (v_{ij} - \bar{v}_j)(v_{ij} - \bar{v}_j)^\top.$$

O quadrado da distância de Mahalanobis entre $\bar{\mathbf{v}}_1$ e $\bar{\mathbf{v}}_2$ é

$$d_m^2 = (\bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_2)^\top \mathbf{S}_c^- (\bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_2), \quad (3.45)$$

em que $\mathbf{S}_c = [n_1 \widehat{\boldsymbol{\Sigma}}_1 + n_2 \widehat{\boldsymbol{\Sigma}}_2] / (n_1 + n_2 - 2)$ e \mathbf{S}_c^- é a inversa generalizada de Moore-Penrose de \mathbf{S}_c . Portanto, a estatística de Hotelling é dada por

$$F_H = \frac{n_1 n_2 (n_1 + n_2 - a - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)a} d_m^2, \quad (3.46)$$

no qual $a = 2k - 2$ é a dimensão do espaço tangente. Ademais, sob H_0 , F_H tem distribuição F-Snedecor com parâmetros $(a, n_1 + n_2 - a - 1)$ no numerador e denominador, respectivamente. A hipótese nula é rejeitada para grandes valores de F_H . É importante ressaltar que, as coordenadas tangentes são calculadas através de uma mesmo polo, ou seja, na forma média geral entre os dois grupos, já que sob H_0 as formas médias são iguais (DRYDEN; MARDIA, 2016, p. 187).

3.6.1.2 Estatística de James

A estatística de James é uma espécie de generalização da estatística F_H . A diferença é que não é utilizada a matriz de covariâncias combinada. Assim, a estatística de James é dada

por

$$F_J = (\bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_2)^\top \left(\frac{1}{n_1} \widehat{\boldsymbol{\Sigma}}_1 + \frac{1}{n_2} \widehat{\boldsymbol{\Sigma}}_2 \right)^{-1} (\bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_2). \quad (3.47)$$

Essa estatística possui sob H_0 , distribuição Qui-Quadrado com a graus de liberdade (χ_a^2) (SEBER, 1984, p. 115).

3.6.1.3 Estatística de Goodall

Essa estatística foi proposta por Goodall (1991) e é baseada no quadrado da distância procrustes completa (2.5). Assim, a estatística de Goodall para comparar duas formas médias é

$$F_G = \frac{n_1 + n_2 - 2}{n_1^{-1} + n_2^{-1}} \frac{d_F^2(\widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\mu}}_2)}{\sum_{i=1}^{n_1} d_F^2(\mathbf{z}_{i1}, \widehat{\boldsymbol{\mu}}_1) + \sum_{i=1}^{n_2} d_F^2(\mathbf{z}_{i2}, \widehat{\boldsymbol{\mu}}_2)}, \quad (3.48)$$

em que $\widehat{\boldsymbol{\mu}}_1$ e $\widehat{\boldsymbol{\mu}}_2$ são as estimativas das formas médias para cada população, utilizando o estimador descrito em (2.10). Sob H_0 , a estatística de Goodall possui distribuição $F_{a, (n_1+n_2-2)a}$.

3.6.2 Testes para Várias Populações

Considere o caso em que existem g grupos de objetos. Seja $j = 1, 2, \dots, g$ o índice dos grupos com tamanho n_j , no qual $n = \sum_{j=1}^g n_j$. A forma média para o j -ésimo grupo é dada por $\boldsymbol{\mu}_j$. Sejam $\mathbf{z}_{1j}, \dots, \mathbf{z}_{n_j j}$ a amostra de pré-formas para o j -ésimo grupo. As hipóteses para testar várias amostras são definidas por

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g$$

$$H_1 : \exists \text{ pelo menos um } i \text{ e } j \text{ tal que } \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j.$$

3.6.2.1 Extensão da Estatística Goodall

A extensão da Estatística de Goodall para várias populações, porém considerando que a amostra é balanceada com $n_j = n$, é dada por

$$F_{G_2} = n(n-1)g \frac{\sum_{j=1}^g d_F^2(\widehat{\boldsymbol{\mu}}_j, \widehat{\boldsymbol{\mu}}_0)}{(g-1) \sum_{j=1}^g \sum_{i=1}^n d_F^2(\mathbf{z}_{ij}, \widehat{\boldsymbol{\mu}}_j)}, \quad (3.49)$$

em que $\widehat{\boldsymbol{\mu}}_0$ é a forma média geral de todas as amostras, ou seja, o autovetor correspondente ao maior autovalor de

$$\mathbf{S}_t = \sum_{j=1}^g \sum_{i=1}^n \mathbf{z}_{ij} \mathbf{z}_{ij}^*.$$

Sob a hipótese nula, F_{G_2} tem aproximadamente distribuição $F_{(g-1)a, g(n-1)a}$ (GOODALL, 1991).

3.6.2.2 Estatística Lambda

A estatística lambda foi proposta por Amaral, Dryden e Wood (2007) e mostrou-se um bom desempenho, em termos do erro do tipo I em testes de hipóteses, do que as estatísticas mencionadas anteriormente, quando existe baixa concentração nos dados. A estatística de teste é dada por

$$\lambda_{\min} = \min_{\boldsymbol{\mu}_0} 2\boldsymbol{\mu}_0^* \widehat{\mathbf{A}}_0 \boldsymbol{\mu}_0, \quad \text{com } \widehat{\mathbf{A}}_0 = n \sum_{j=1}^g \widehat{\mathbf{M}}_j^* \widehat{\mathbf{G}}_j^{-1} \widehat{\mathbf{M}}_j, \quad (3.50)$$

em que $\widehat{\mathbf{M}}_j = [\mathbf{u}_2^j, \mathbf{u}_3^j, \dots, \mathbf{u}_{k-1}^j]^*$ é a matriz de dimensão $(k-2) \times (k-1)$, na qual as suas linhas são formadas pelos autovetores (excluindo o primeiro autovetor) da matriz soma produto \mathbf{S}_j , definida em (2.9), para a j -ésima população. Considere $\lambda_1^j, \lambda_2^j, \dots, \lambda_{k-1}^j$ como sendo os autovalores de \mathbf{S}_j , assim, $\widehat{\mathbf{G}}_j = [G_{u,v}]_{u,v=2}^{k-1}$ no qual

$$\widehat{G}_{u,v} = n^{-1} (\lambda_1^j - \lambda_u^j)^{-1} (\lambda_1^j - \lambda_v^j)^{-1} \sum_{i=1}^{n_j} (\mathbf{u}_u^{j*} \mathbf{z}_{ij}) (\mathbf{z}_{ij}^* \mathbf{u}_v^j) (\mathbf{z}_{ij}^* \widehat{\boldsymbol{\mu}}_j) (\widehat{\boldsymbol{\mu}}_j^* \mathbf{z}_{ij}),$$

em que $\widehat{\boldsymbol{\mu}}_j = \mathbf{u}_1^j$ é a estimativa da forma média para o j -ésimo grupo. Sob H_0 , a distribuição de λ_{\min} é $\chi_{2(g-1)(k-2)}^2$.

3.6.2.3 Estatísticas de Teste no Plano Real

Na Estatística, o procedimento utilizado para comparar médias entre várias populações é a análise de variância multivariada ou MANOVA (do inglês *multivariate analysis of variance*). Esse método é uma extensão da análise de variância, mais conhecido como ANOVA (*analysis of variance*). Suponha que temos as amostras $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \dots, \mathbf{x}_{n_jj}$, com $j = 1, 2, \dots, g$ e n_j sendo o número de amostras para o j -ésimo grupo. Considere ainda que

$$\bar{\mathbf{x}}_j = \frac{\sum_{i=1}^{n_j} \mathbf{x}_{ij}}{n_j},$$

$$\bar{\mathbf{x}} = \frac{\sum_{j=1}^g \sum_{i=1}^{n_j} \mathbf{x}_{ij}}{n},$$

no qual $n = \sum_{j=1}^g n_j$. Assim sendo, desejamos testar se as médias entre os g grupos são iguais ou pelo menos uma média é diferente das demais através da MANOVA. Formalmente, resumimos os cálculos que são utilizados no cálculo das estatísticas de teste na Tabela 4.

Tabela 4 – Exemplos de alguns kernels complexos.

Fonte de Variação	Soma de Quadrados	Graus de Liberdade
Tratamento	$\mathbf{B} = \sum_{j=1}^g n_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^\top$	$\nu_{\mathbf{H}} = g - 1$
Residual	$\mathbf{E} = \sum_{j=1}^g \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)^\top$	$\nu_{\mathbf{E}} = n - g$
Total	$\mathbf{B} + \mathbf{E} = \sum_{j=1}^g \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})^\top$	$n - 1$

Sejam $\lambda_1^\dagger, \lambda_2^\dagger, \dots, \lambda_s^\dagger$ os autovalores de $\mathbf{B}\mathbf{E}^{-1}$, com $s = \min\{\nu_{\mathbf{H}}, p\}$. As estatísticas de teste são baseadas nos autovalores de $\mathbf{B}\mathbf{E}^{-1}$ e são dadas por

$$\text{Wilks: } \Lambda_{Wilks} = \prod_{i=1}^s \frac{1}{1 + \lambda_i^\dagger}, \quad (3.51)$$

$$\text{Roy: } \Lambda_{Roy} = \lambda_1^\dagger, \quad (3.52)$$

$$\text{Lawley-Hotelling: } \Lambda_{LH} = \sum_{i=1}^s \lambda_i^\dagger, \quad (3.53)$$

$$\text{Pillai: } \Lambda_{Pillai} = \sum_{i=1}^s \frac{\lambda_i^\dagger}{1 + \lambda_i^\dagger}. \quad (3.54)$$

Os testes não são equivalentes, ou seja, em uma dada amostra, podem levar conclusões diferentes. Contudo, para amostras grandes, todas essas estatísticas são equivalentes (JOHNSON; WICHERN, 2007, p. 303). A distribuição para cada estatística pode ser muito complicada e geralmente é feita uma modificação para que se tenha uma distribuição F -Snedecor. Porém, não existe uma forma geral para todos os casos, pois a modificação irá depender do número de variáveis e de grupos. Entretanto, existem algumas aproximações que estão descritas em Rencher e Schaalje (2008, p. 163-168) e elas são dadas por

$$\text{Wilks: } F_{Wilks} = \frac{1 - \Lambda_{Wilks}^{1/t}}{\Lambda_{Wilks}^{1/t}} \frac{\mathbf{df}_2}{\mathbf{df}_1}, \quad (3.55)$$

em que \mathbf{df}_1 e \mathbf{df}_2 são os graus de liberdade do numerador e denominador da distribuição F com

$$\mathbf{df}_1 = p\nu_{\mathbf{H}}, \quad \mathbf{df}_2 = wt - \frac{1}{2}(p\nu_{\mathbf{H}} - 2),$$

$$w = \nu_{\mathbf{E}} + \nu_{\mathbf{H}} - \frac{1}{2}(p + \nu_{\mathbf{H}} + 1), \quad t = \sqrt{\frac{p^2\nu_{\mathbf{H}}^2 - 4}{p^2 + \nu_{\mathbf{H}}^2 - 5}}.$$

$$\text{Roy: } F_{\text{Roy}} = \frac{(v_{\mathbf{E}} - a - 1)\lambda_1^\dagger}{a}, \quad (3.56)$$

com a graus de liberdade do denominador e $v_{\mathbf{E}} - a - 1$ no denominador, em que $a = \max\{p, v_{\mathbf{E}}\}$.

$$\text{Lawley-Hotelling: } F_{\text{LH}} = \frac{2(sr + 1)\Lambda_{\text{LH}}}{s^2(2m + s + 1)}, \quad (3.57)$$

com $m = \frac{1}{2}(|v_{\mathbf{H}} - p| - 1)$ e $r = \frac{1}{2}(v_{\mathbf{E}} - p - 1)$, que possui aproximadamente distribuição F com os parâmetros $s(2m + s + 1)$ e $2(sr + 1)$.

$$\text{Pillai: } F_{\text{Pillai}} = \frac{(2r + s + 1)\Lambda_{\text{Pillai}}}{(2m + s + 1)(s - \Lambda_{\text{Pillai}})}, \quad (3.58)$$

que é aproximadamente distribuída como $F_{s(2m+s+1), s(2r+s+1)}$. A hipótese nula é rejeitada para todo os casos, quando cada estatística é maior que o quantil da distribuição F correspondente.

Como estamos trabalhando com as pré-formas, não é possível utilizar essas estatísticas no plano real. Assim, serão utilizadas as CT (2.11) das pré-formas.

Um detalhe importante, é o fato de que a estatística λ_{\min} também pode ser utilizada para o caso em que $g = 2$ (AMARAL; DRYDEN; WOOD, 2007). Entretanto, o mesmo não acontece para os testes da MANOVA, pois se $g = 2$, implica que $v_H = 1$. Conseqüentemente, $s = \min\{v_H, p\} = 1$, tornando as estatísticas (3.51), (3.52), (3.53) e (3.54) equivalentes, pois elas dependerão apenas de λ_1^\dagger . Além disso, existe uma relação entre a estatística de Hotelling (3.45) e Wilks (3.51), assim, tornando-se equivalente às demais estatísticas (MARDIA; KENT; BIBBY, 1979, p. 138-140). No Algoritmo 7, estão expostos os passos para a montagem e classificação.

Algoritmo 7: Algoritmo *Hill-Climbing*.

Entrada: Sejam $\mathbf{z}_{1j}, \mathbf{z}_{2j}, \dots, \mathbf{z}_{nj}$, com $j = 1, 2, \dots, g$ as amostras de cada população.

Saída: Classificador *Hill-Climbing*.

- 1:** As amostras são utilizadas para definir o valor de referência da estatística teste;
 - 2:** Seja \mathbf{z}' uma pré-forma que se deseja classificar;
 - 3:** Coloque o objeto \mathbf{z}' em todos os grupos considerados e verifique qual grupo forneceu o maior valor da estatística escolhida. Para o caso de dois grupos, considere as estatísticas (3.46), (3.47), (3.48) e (3.50). Se existir g grupos ($g > 2$), as estatísticas (3.49), (3.55), (3.56), (3.57), (3.58) e (3.50) são utilizadas;
 - 4:** O objeto \mathbf{z}' é classificado no grupo que forneceu o maior valor da estatística escolhida no passo anterior.
-

3.7 MÉTODOS DE AGRUPAMENTO

Os métodos de agrupamento referem-se a um conjunto amplo de técnicas cujo o objetivo é formar subgrupos ou grupos em um conjunto de dados. Para formar esses grupos, procura-se dividir o conjunto de dados através da similaridade, ou seja, as observações são bem semelhantes dentro de cada grupo, enquanto que as observações em diferentes grupos são bastantes diferentes (JAMES *et al.*, 2013, p. 385). Por exemplo, suponha que temos um conjunto de n observações com p características (variáveis), em que essas observações podem corresponder a amostras de clientes em uma loja. As p características podem ser a renda, idade, valor gasto, número de filhos, etc. Com isso, poderíamos formar grupos de clientes e direcionar determinadas ofertas levando em conta essas variáveis.

Existem vários métodos de agrupamento, mas o nosso foco é no método chamado de k-médias (KM) e em uma extensão chamada de kernel k-médias (KKM). O Procrustes kernel k-médias (PKKM) foi proposto por Jayasumana *et al.* (2013) para as pré-formas, em que mostrou-se bem superior ao KM.

3.7.1 K-médias

O KM é um método simples, em que ele consiste em alocar os objetos em g grupos, $C = (C_1, C_2, \dots, C_g)$. Inicialmente, para utilizar o método é necessário fornecer o número de grupos g , então o algoritmo irá atribuir todas as observações em um dos g grupos. Formalmente, suponha que temos uma amostra $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ contendo n observações de uma variável aleatória p -dimensional \mathbf{Z} . Considere $\boldsymbol{\mu}_j$ como sendo a média (forma média) do j -ésimo grupo, com $j = 1, 2, \dots, g$. Nosso objetivo é alocar os objetos para cada grupo, com o intuito de minimizar a soma de quadrados dos erros sobre todos grupos, ou seja,

$$J_1 = \sum_{j=1}^g \sum_{i \in C_j} d^2(\mathbf{z}_i, \boldsymbol{\mu}_j), \quad (3.59)$$

no qual $d(\cdot, \cdot)$ é alguma das distâncias (2.5), (2.6) ou (2.7). O Algoritmo 8 descreve os passos para a formação dos grupos.

Algoritmo 8: Método k-médias

Entrada: Conjunto de pré-formas $\{\mathbf{z}_i\}_{i=1}^n$, número de grupos g e as médias de cada grupo.

Saída: Grupos C_1, C_2, \dots, C_g .

- 1:** Calcular a distância de cada observação para a forma média utilizando alguma das distâncias (2.5), (2.6) ou (2.7);
 - 2:** Alocar o objeto para o grupo que tiver a menor distância;
 - 3:** Recalcular a média para cada grupo;
 - 4:** Repetir os passos de 1 a 3 até que não aconteça uma redução em (3.59), ou ainda, quando o algoritmo atingir um número máximo de iterações predeterminada.
-

Geralmente, não temos os centroides (médias) para aplicar o KM. O que pode ser feito, é selecionar aleatoriamente g observações e utilizá-las como os centroides iniciais. Uma outra maneira, seria selecionar aleatoriamente g observações em um número grande de vezes, por exemplo: 100 vezes, e aplicar o KM em cada uma dessas situações. Por fim, escolhe-se como centroides iniciais o conjunto de observações que proporcionou um menor valor de (3.59). Evidentemente, o método também é aplicado em vetores reais, porém é necessário mudar a distância, por exemplo, a distância euclidiana ou a de Mahalanobis. O Algoritmo 9 expõe os detalhes para conseguir a partição inicial. Para mais detalhes sobre esse método, ver Bishop (2006, p. 424-425).

Algoritmo 9: Partição Inicial.

Entrada: Uma amostra de pré-formas $\mathbf{z}_1, \dots, \mathbf{z}_n$, número de iterações N e o número de grupos g .

Saída: Centroides iniciais para o agrupamento.

- 1:** Sejam $\mathbf{z}_1, \dots, \mathbf{z}_n$ as pré-formas de uma amostra aleatória de objetos.
 - 2:** Gere g números aleatórios sem reposição de $1, \dots, n$ para serem os índices.
 - 3:** Seja C_j o centro do j -ésimo grupo. Faça $C_j = \mathbf{z}_j$, ($j = 1, \dots, g$), no qual j é o índice escolhido no passo 2;
 - 4:** Aloque \mathbf{z}_l , em que $l = 1, \dots, n$, para o centro de grupo C_j , mais próximo utilizando alguma das distâncias (2.5), (2.6) ou (2.7) como critério;
 - 5:** Repetir N vezes os passos de 2 a 4;
 - 6:** Escolha a partição que forneceu o menor valor de (3.59).
-

3.7.1.1 Kernel k-médias

O KKM é uma extensão do método KM, no qual realiza o agrupamento em um espaço não linear. Esse método foi proposto por Schölkopf, Smola e Müller (1996) e a ideia é acrescentar um kernel, como aquele visto na SVM, na equação (3.59). Vimos que o kernel é essencial para a SVM, proporcionando um melhor desempenho que o kernel linear, portanto, é esperado que o mesmo aconteça com a adição do kernel no KM. Assim, iremos utilizar o kernel Procrustes Gaussiano que foi apresentado por Jayasumana *et al.* (2013), em que nele é utilizado a distância Procrustes completa (2.5) e que possui a seguinte expressão

$$K(\mathbf{z}, \mathbf{z}') = \exp\left(-\frac{d_F^2(\mathbf{z}, \mathbf{z}')}{2h_2}\right) = \exp\left(-\frac{1 - |\mathbf{z}, \mathbf{z}'|^2}{2h_2}\right), \quad (3.60)$$

no qual $h_2 \in \mathbb{R}^+$ é um parâmetro de suavização. As outras distâncias (2.6) e (2.7) não satisfazem as condições do Teorema 3, por isso, não podem ser consideradas um kernel (JAYASUMANA *et al.*, 2013). O próximo passo é escolher a distância para formular a minimização dos erros. Oliveira (2016) propôs utilizar a própria distância procrustes completa na função objetivo (3.59). Assim, a equação (3.59) é modificada para

$$J_2 = \sum_{j=1}^g \sum_{i \in C_j} (1 - |\phi(\mathbf{z}_i^*) \boldsymbol{\mu}_j^\phi|^2), \quad (3.61)$$

em que, o valor do j -ésimo vetor de médias é

$$\boldsymbol{\mu}_j^\phi = \sum_{\mathbf{z} \in C_j} \frac{\phi(\mathbf{z})}{\#C_j}, \quad (3.62)$$

no qual o operador # indica o número de elementos do grupo j . Como a função $\phi(\cdot)$ não possui uma forma explícita, não é possível calcular os vetores de médias em (3.62) de forma clara. Porém, podemos expandir a equação (3.61) utilizando o fato de que

$$\begin{aligned}
|\phi(\mathbf{z}_i^*)\boldsymbol{\mu}_j^\phi|^2 &= \left| \phi(\mathbf{z}_i^*) \sum_{\mathbf{z} \in C_j} \frac{\phi(\mathbf{z})}{\#C_j} \right|^2 \\
&= \left| \sum_{\mathbf{z} \in C_j} \phi(\mathbf{z}_i^*) \frac{\phi(\mathbf{z})}{\#C_j} \right|^2 \\
&= \left| \sum_{\mathbf{z} \in C_j} \frac{K(\mathbf{z}_i, \mathbf{z})}{\#C_j} \right|^2 \\
&= \left(\sum_{\mathbf{z} \in C_j} \frac{K(\mathbf{z}_i, \mathbf{z})}{\#C_j} \right)^2.
\end{aligned} \tag{3.63}$$

Assim sendo, a equação (3.61) pode ser reescrita para

$$J_2 = \sum_{j=1}^g \sum_{i \in C_j} \left[1 - \left(\sum_{\mathbf{z} \in C_j} \frac{K(\mathbf{z}_i, \mathbf{z})}{\#C_j} \right)^2 \right].$$

Note que o método PKKM não possui um etapa para atualização dos vetores de médias, como no caso do KM, pelo fato do mapeamento $\phi(\cdot)$ não possui uma forma explícita. Porém, mesmo sem esse mapeamento é possível realizar a alocação das observações nos grupos via equação (3.63) na distância $d_F^2(\cdot, \cdot)$. O mecanismo de alocação é o mesmo do método KM. Os objetos são alocados para cada grupo em cada iteração, até que não haja nenhuma redução em (3.61) ou até que se chegue em um número máximo de iterações. O Algoritmo 10 resume os passos para a utilização do PKKM.

Algoritmo 10: Método Procrustes kernel k-médias.

Entrada: Conjunto de pré-formas $\{\mathbf{z}_i\}_{i=1}^n$, número de grupos g e a alocação inicial.

Saída: Grupos C_1, C_2, \dots, C_g .

1: Calcular todos os valores de $K(\mathbf{z}_i, \mathbf{z}_j)$ (3.60), para $i, j = 1, 2, \dots, n$;

2: Calcular todas as distâncias de cada observação para cada grupo através de

$$\sum_{j \in C_i} \left[1 - \left(\sum_{\mathbf{z} \in C_i} \frac{K(\mathbf{z}_j^*, \mathbf{z})}{\#C_i} \right)^2 \right];$$

3: Alocar em cada grupo, as observações que obtiveram menor valor da distância;

4: Repetir o passo 2 e 3 até que não aconteça uma redução em (3.61), ou ainda, se o algoritmo atingir um número máximo de iterações predeterminada.

3.7.1.2 Classificação Por Meio de Métodos de Agrupamento

Os métodos de agrupamento podem ser adaptados para serem classificadores. Nesse caso, os rótulos dos dados de treinamento já são conhecidos, não havendo a necessidade de aplicar os métodos de agrupamento. Para classificar uma observação em um dos grupos, basta utilizar o critério adotado para o agrupamento, ou seja, é calculada a distância da observação para os centroides e a observação é alocada para o grupo que produziu a menor distância. Entretanto, essa ideia necessita ser melhorada, pois se existir uma grande variabilidade entre os dados, a chance de cometer um erro é muito grande. Assim, Hastie, Tibshirani e Friedman (2001, p. 460) recomendam que o método de agrupamento seja aplicado em cada grupo nos dados de treinamento, produzindo R centroides em cada grupo e utilizar esses centroides para a classificação. Essa abordagem é conhecida como método protótipo e será utilizada para a classificação. Esse classificador será chamado de Procrustes kernel k-médias com protótipo ($PKMM_p$) e o Algoritmo 11 mostra como esse classificador é utilizado.

Algoritmo 11: Método Protótipo baseado no algoritmo Procrustes kernel k-médias.

Entrada: Conjunto de treinamento $\{\mathbf{z}_{ij}\}_{i=1}^{n_j}$, com $j = 1, 2, \dots, g$ e o número de centroides R .

Saída: Classificador Procrustes kernel k-médias com protótipo.

- 1:** Utilize o Algoritmo 9 da partição inicial, impondo R centroides para o j -ésimo grupo, com $j = 1, 2, \dots, g$;
 - 2:** Seja \mathbf{z}' uma nova observação e C_{jr} o r -ésimo centroide do j -ésimo grupo. Calcule as distâncias da nova observação para todos os centroides em todos os grupos;
 - 3:** Aloque \mathbf{z}' para o grupo que produziu a menor distância.
-

3.8 MÉTODOS *ENSEMBLE*

Os métodos *ensemble* (ME) são utilizados para combinar classificadores, com o objetivo de melhorar os resultados. Segundo Zhou (2012, p. 16) existem 3 tópicos de contribuições iniciais que levaram à área atual de ME.

- Combinação de Classificadores

Neste tópico, o objetivo é trabalhar com classificadores fortes, ou seja, geralmente possuem um bom desempenho. Assim sendo, tenta-se criar regras de combinação para obter classificadores combinados mais fortes;

- Conjunto de Fracos Aprendizes

Neste segmento, o objetivo é trabalhar com modelos (regressão, classificador agrupamento) fracos e tentar projetar algoritmos poderosos para aumentar o desempenho de fraco para forte. Este tópico deu origem aos métodos *Boosting* e *Bagging*;

- Mistura de Especialistas

Esse tópico foi estudado principalmente pela comunidade de redes neurais. Geralmente, é considerado uma estratégia de dividir e conquistar, tentar aprender uma mistura de modelos paramétricos em conjunto e usar regras de combinação para obter uma solução global.

Zhou (2012, p 17-18) menciona várias aplicações dos ME, dentre elas, a *KDD-Cup* (uma famosa competição de mineração de dados) e a *Netflix Prize*. Na *KDD-Cup* entre os anos de 2009 a 2011, todos os vencedores do primeiro e segundo lugares utilizaram ME. No *Netflix Prize* o objetivo era melhorar a precisão das previsões sobre o quanto alguém vai gostar de um filme baseado nas preferências. Caso alguma equipe conseguisse melhorar em 10% a precisão do próprio algoritmo da *Netflix*, ganharia o prêmio de um milhão de dólares. O time vencedor propôs uma solução baseada na combinação de vários classificadores.

O interesse nessa seção para a presente pesquisa, será o tópico de combinação de classificadores. A seguir, são descritas algumas maneiras para obter essa combinação.

3.8.1 Combinação de Classificadores

A combinação depende da saída de cada modelo ou classificador, nesse caso, para os classificadores mostrados anteriormente, a saída é justamente o rótulo do grupo. Considere um conjunto de T classificadores individuais $\{F_1, F_2, \dots, F_T\}$ e o objetivo é combinar as saídas de cada classificador de um conjunto de g possíveis grupos $\{c_1, c_2, \dots, c_g\}$. Para cada observação \mathbf{x} , as saídas do classificador F_i são dadas como um vetor g -dimensional $(F_i^1(\mathbf{x}), F_i^2(\mathbf{x}), \dots, F_i^g(\mathbf{x}))$, no qual $F_i^j(\mathbf{x})$ é a saída do classificador i para o grupo c_j . Ademais, $F_i^j(\mathbf{x}) \in \{0, 1\}$, em que $F_i^j(\mathbf{x})$ assume o valor um se F_i^j predizer o grupo c_j e zero caso contrário. A seguir, pode-se definir os seguintes métodos de combinação.

3.8.1.1 Maioria dos Votos

Esse método funciona da mesma maneira como em muitas democracias no mundo inteiro para eleger um presidente da república, no qual é necessário a maioria dos votos válidos.

Cada classificador vota em um dos rótulos e o rótulo de saída final é aquele que recebe mais da metade dos votos. Se nenhum dos rótulos receber mais da metade dos votos, será dada uma opção de rejeição e o classificador combinado não faz previsão (ZHOU, 2012, p. 72). Isto é, a saída para o classificador combinado é

$$F_{comb}(\mathbf{x}) = \begin{cases} c_j, & \text{se } \sum_{i=1}^T F_i^j(\mathbf{x}) > \frac{1}{2} \sum_{r=1}^g \sum_{i=1}^T F_i^r(\mathbf{x}) \\ \text{não classifica,} & \text{caso contrário.} \end{cases}$$

3.8.1.2 Voto de Pluralidade

Diferentemente do voto majoritário, em que é necessário a maioria absoluta dos votos para classificar em um dos grupos, o voto de pluralidade é aquele que apenas necessita da maioria simples dos votos, ou seja, a classificação é feita para o grupo que receber o maior número de votos (ZHOU, 2012, p. 73). Assim sendo, o classificador combinado fica na forma

$$F_{comb}(\mathbf{x}) = c_j, \quad j = \arg \max_j \sum_{i=1}^T F_i^j(\mathbf{x}).$$

Evidentemente, o voto de pluralidade não tem uma opção de rejeição, porém pode acontecer empate caso o número de classificadores T seja um número par. Além disso, para o caso de dois grupos, o voto de pluralidade coincide com a votação majoritária (ZHOU, 2012, p. 73).

3.8.1.3 Votação Ponderada

Caso os classificadores possuam desempenhos diferentes, é razoável atribuir pesos diferentes na rotulação de cada classificador. Com esse intuito, o classificador combinado pode ser visto como

$$F_{comb}(\mathbf{x}) = c_j, \quad j = \arg \max_j \sum_{i=1}^T w_i F_i^j(\mathbf{x}),$$

em que w_i é o peso dado ao i -ésimo classificador (ZHOU, 2012, p. 74). Além disso, seria interessante que essa combinação de pesos maximizasse a probabilidade de acerto do classificador combinado. Dessa forma, considerando uma combinação de T classificadores $\{F_1, F_2, \dots, F_T\}$ independentes, com acurácias individuais p_1, \dots, p_T . O valor de w_i que maximiza a probabilidade

de acerto satisfaz

$$w_i \propto \log \frac{p_i}{1 - p_i}, \quad i = 1, 2, \dots, T, \quad (3.64)$$

no qual o símbolo \propto é o símbolo de proporcionalidade, ou seja, os pesos devem ser proporcionais ao desempenho dos classificadores individuais (KUNCHEVA, 2004, p. 124-125). Vale ressaltar, que o peso obtido em (3.64) é obtido impondo independência dos classificadores, porém isso não acontece, pois os classificadores são treinados sobre o mesmo problema e eles são, geralmente, altamente correlacionados (ZHOU, 2012, p. 75) e nem sempre leva a um desempenho melhor que a votação por maioria.

3.9 VALIDAÇÃO CRUZADA

Quando se tem grandes amostras, é possível separar uma parte delas para averiguar o quão bom é o classificador ou um modelo. Geralmente, o número de observações na análise de forma não é muito grande, impossibilitando essa abordagem. Com o intuito de resolver esse problema, pode-se utilizar os métodos de validação cruzada. Dentre eles, os mais utilizados são o *leave-one-out* e o *K-Fold*. No método *K-fold*, os dados são divididos aleatoriamente em K partes aproximadamente iguais. A j -ésima parte é utilizada para a predição e as $K - 1$ partes restantes, para o treinamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2001, p. 242). Como ilustração, considere $K = 5$ e o cenário é descrito na Figura 13.

1	2	3	4	5
Treinamento	Teste	Treinamento	Treinamento	Treinamento

**Figura 13 – Esquema da validação cruzada utilizando o método $K - fold$ com $K = 5$.
Fonte: Hastie, Tibshirani e Friedman (2001, p. 242)**

Para a parte referente a $j = 2$ (segunda caixa na Figura (13)), o classificador é ajustado utilizando os dados das caixas 1, 3, 4 e 5, em que os dados na caixa 2 são utilizados para a classificação (predição) e o erro é calculado para esse conjunto de teste. Esse procedimento é feito para todas as partes, ou seja, $j = 1, 2, \dots, 5$ e no final os percentuais estimados dos erros são combinados, gerando uma estimativa geral para a taxa de erro do classificador.

O método *leave-one-out* é um caso particular do método *K-fold* e é frequentemente utilizado quando se tem poucos dados. Nessa situação, o número de partes é igual ao número de observações, ou seja, $K = n$. Assim, a j -ésima observação ($j = 1, 2, \dots, n$) é excluída e utilizada para a predição, enquanto que as $n - 1$ observações são utilizadas para o treinamento. Esse processo é repetido n vezes, no final, calcula-se o percentual de predições erradas.

No Algoritmo 12 estão descritos os passos para a utilização do método da validação cruzada.

Algoritmo 12: Validação Cruzada.

Entrada: Conjunto de treinamento $T = \{\mathbf{x}_i, y_i\}_{i=1}^n$ e o número de conjuntos K .

Saída: Taxa de erro.

- 1: Separe as n amostras aleatoriamente em K conjuntos com tamanhos aproximadamente iguais;
- 2: Retire o j -ésimo conjunto e utilize os $n - 1$ conjuntos restantes para o treinamento do classificador;
- 3: Utilize o j -ésimo conjunto para a predição e calcule o número de erros e_j ;
- 4: Repita os passos 2 e 3 para $j = 1, 2, \dots, K$;
- 5: Calcule a taxa de erro dada por

$$VC_{erro} = \frac{1}{K} \sum_{j=1}^K e_j.$$

3.10 SIMULAÇÃO

Com o intuito de avaliar o desempenho dos classificadores, foi realizado um experimento de simulação no espaço dos marcos com diferentes variabilidades. São considerados dois quadrados com oito marcos, sendo a forma média em cada grupo, como mostrado na Figura 14. Para simular os dados, cada marco é visto como sendo o vetor de médias de uma variável aleatória tendo distribuição normal bivariada com matriz de covariância $\Sigma = \sigma^2 I_2$, ou seja, são geradas n amostras considerando oito variáveis aleatórias normais bivariada independentes para cada um dos dois grupos para diversos valores de σ^2 . O Algoritmo 13 descreve os passos para a geração de amostras para o espaço dos marcos.

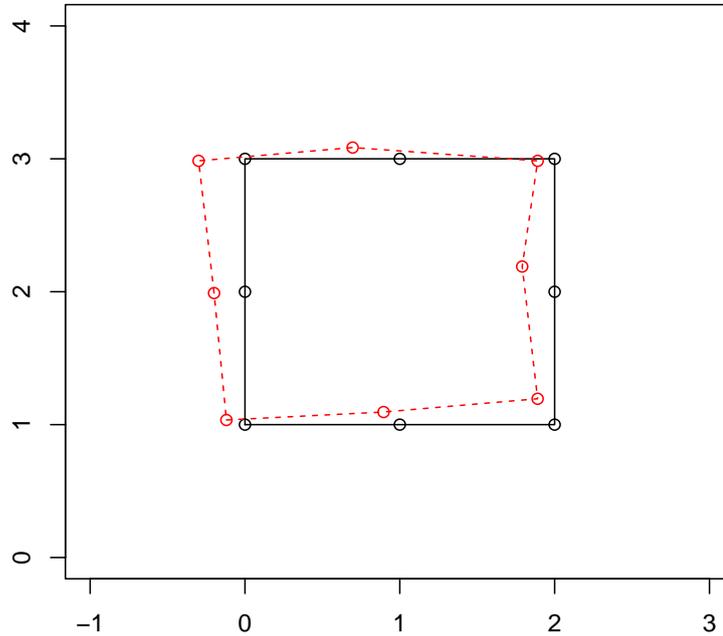


Figura 14 – Formas médias para os dois grupos.

Algoritmo 13: Simulação.

Entrada: Tamanho da amostra n , número de marcos k , matriz de configuração $\tilde{\mathbf{Y}}_{k \times 2}$ representando a figura geométrica de interesse e o valor de σ^2 .

Saída: Matrizes de configuração $\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \dots, \tilde{\mathbf{Y}}_n$.

1: Faça $\boldsymbol{\mu} = (\tilde{y}_{11}, \tilde{y}_{12}, \tilde{y}_{21}, \tilde{y}_{22}, \dots, \tilde{y}_{k1}, \tilde{y}_{k2})^\top$;

2: Gere n amostras aleatórias de $\mathbf{X}_j \sim N_{2k}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($j = 1, 2, \dots, n$), em que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_k \end{pmatrix},$$

no qual $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{(2k)j})^\top$ é o vetor da j -ésima realização. **3:** Faça cada par de observações $(x_{1j}, x_{2j}), (x_{3j}, x_{4j}), \dots, (x_{(2k-1)j}, x_{(2k)j})$ como sendo as linhas de $\tilde{\mathbf{Y}}_j$, com $j = 1, 2, \dots, n$.

O estudo de simulação considerou todos os classificadores com diferentes configurações utilizando a validação cruzada com $n = 100$ amostras (50 para cada grupo) e $K = 5$. Foi considerado também, vários valores de $\sigma^2 = 0,001; 0,005; 0,009; 0,01; 0,014; 0,018; 0,02$. As configurações para cada classificador são:

- SVM Complexa: $C = 1, 10, 10^2, 10^3, 10^4, 10^5, 10^6$, para os kernels
 - Polinomial Complexo: $a_1 = 1, a_2 = 0$ e $a_3 = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ (o kernel linear acontece quando $a_3 = 1$).
 - Gaussiano Complexo: $\sigma_r^2 = 0,01; 0,05; 0,1; 0,5; 1; 10; 10^2; 10^3; 10^4$.
- SVM Complexa com Pesos: Foram utilizadas as formas médias da Figura 14. Posteriormente, foram utilizadas as distâncias (2.5), (2.6) e (2.7) para calcular os pesos (3.42). Além disso, foi utilizada a mesma configuração para os kernels como na SVM complexa.
- Procrustes kernel K-Médias: O PKKM foi utilizado com $h_1 = 0, 1; 0,5; 1; 1,5; 2; 2,5; 3; 3,5; 4$. O $PKKM_p$ foi utilizado com os mesmos valores de h_1 com os seguintes números de centroides $R = 2, 3, 4, 5$.
- Bayes: O classificador de Bayes foi utilizado com as distribuições Bingham e Watson complexa.
- Estimação de Densidade: O classificador Bayes-ED foi utilizado com as distribuições normal multivariada complexa para os valores de suavização $h_2 = 0,005; 0,01; 0,05; 0,1; 0,5; 1; 1,5; 2$. Ademais, foi utilizado a distribuição t-Student multivariada complexa com a mesma configuração, utilizado os graus de liberdade de 1 a 10.
- *Hill-Climbing*: O HC foi utilizado para as estatísticas de James, Hotteling, Goodall e Lambda.
- *Ensemble*: Foram utilizados os três métodos de combinação de votos, o voto majoritário (VM), o voto de pluralidade (VP) e o voto de pluralidade ponderado (VPP) para os melhores classificadores em relação a taxa de erro em cada situação.

Os resultados das simulações estão na Tabela 5. Pode-se concluir que à medida que a variabilidade (σ^2) aumenta, o desempenho tende a piorar para todos os classificadores. De forma geral, a SVM e o classificador de Bayes-ED produziram os melhores resultados, praticamente em todas as situações. A seguir, estão as conclusões para cada classe de classificadores.

- SVM Complexa: O classificador SVM não diferiu com a sua versão com os pesos, praticamente o mesmo desempenho em todos os casos.
- Procrustes kernel K-Médias: O $PKKM_p$ foi superior ao PKKM em praticamente todos os casos.

- Bayes: O discriminante Watson foi melhor nas situações em que a variabilidade é pequena.
- Estimaco de Densidade: A distribuico t-Student multivariada complexa teve um desempenho um pouco melhor do que a distribuico normal.
- *Hill-Climbing*: No houve nenhuma estatística que foi melhor em todos os casos, o resultado varia de acordo com a variabilidade dos dados.
- *Ensemble*: Nesse método, foram utilizados os classificadores $SVM_{\mathbb{C}}(d_F)$ com o kernel polinomial, $PKKM_p$, a distribuico t-Student, o discriminante Watson e a estatística de James. O voto de pluralidade ponderado (VPP) foi o que produziu os melhores resultados em comparaco com as outras combinaçes.

Analisando os classificadores nas suas respectivas classes, percebe-se que os desempenhos da SVM complexa tradicional e SVM com os pesos, so praticamente os mesmos, porm h diferenas entre os kernels. Os kernels polinomial e gaussiano, sempre produziram resultados melhores ou iguais ao kernel linear. Em relao aos classificadores baseados em agrupamento, o $PKKM$ e o $PKKM_p$, percebe-se que na grande maioria dos casos, o $PKKM_p$ obteve um desempenho melhor do que o $PKKM$.

Os classificadores na classe Bayes-ED, praticamente, resultaram no mesmo desempenho entre o uso da normal e t-Student multivariada complexa. No obstante, esses classificadores obtiveram um desempenho melhor em comparaco com o classificador de Bayes paramtrico. Ademais, quando a variabilidade aumenta, o classificador de Bayes utilizando a Watson complexa, mostrou-se melhor em relao a Bingham complexa.

No classificador HC, nota-se que o desempenho das estatísticas de teste so similares, porm a estatística de James piora quando a variabilidade dos dados aumenta, em comparaco com as demais estatísticas. Por fim, observe-se que o método *ensemble* com a utilizao da VPP, obteve resultados melhores do que os outros métodos de combinao. Contudo, no produziu um resultado melhor do que a SVM complexa, porm houve uma melhora na taxa de erro comparada com vrios classificadores individuais.

Tabela 5 – Taxas de erro para a simulação, considerando diferentes configurações de variabilidade.

Classificador	Tipo	σ^2						
		0,001	0,005	0,009	0,01	0,014	0,018	0,02
SVM _C	L	0,0000	0,0100	0,0500	0,0500	0,1000	0,1000	0,1400
	P	0,0000	0,0100	0,0400	0,0400	0,0900	0,0900	0,1500
	G	0,0000	0,0100	0,0400	0,0400	0,0900	0,0900	0,1400
SVM _C (d_F)	L	0,0000	0,0100	0,0500	0,0500	0,1000	0,1100	0,1500
	P	0,0000	0,0100	0,0400	0,0400	0,0900	0,0900	0,1500
	G	0,0000	0,0100	0,0400	0,0500	0,1000	0,1000	0,1400
SVM _C (d_P)	L	0,0000	0,0100	0,0500	0,0500	0,0900	0,1000	0,1500
	P	0,0000	0,0200	0,0400	0,0400	0,0900	0,0900	0,1600
	G	0,0000	0,0100	0,0400	0,0400	0,0900	0,0900	0,1400
SVM _C (ρ)	L	0,0000	0,0100	0,0500	0,0600	0,0900	0,0900	0,1600
	P	0,0000	0,0200	0,0400	0,0500	0,0900	0,0900	0,1500
	G	0,0000	0,0100	0,0400	0,0400	0,0900	0,0900	0,1500
Kernel k-médias	PKKM	0,0000	0,0600	0,1100	0,1000	0,1700	0,2100	0,2000
	PKKM _p	0,0000	0,0400	0,0900	0,1300	0,1400	0,1800	0,2000
Bayes-ED	Normal	0,0000	0,0300	0,0700	0,0600	0,1100	0,0800	0,1400
	t-Student	0,0000	0,0300	0,0600	0,0500	0,1100	0,0800	0,1400
Bayes	Bingham	0,0000	0,0600	0,1500	0,1500	0,2400	0,1700	0,2700
	Watson	0,0000	0,0600	0,1100	0,1100	0,1400	0,1900	0,2000
Hill-Climbing	James	0,0000	0,0500	0,0900	0,1200	0,2100	0,2100	0,2600
	Hotelling	0,0000	0,0700	0,0800	0,1200	0,1700	0,2200	0,2700
	Goodal	0,0000	0,0600	0,1200	0,1100	0,1400	0,1900	0,2000
	Lambda	0,0000	0,0700	0,1500	0,1300	0,1800	0,1800	0,2100
Ensemble	VM	0,0000	0,0600	0,1200	0,1300	0,1500	0,2200	0,2400
	VP	0,0000	0,0300	0,0700	0,1100	0,1200	0,1300	0,1700
	VPP	0,0000	0,0200	0,0500	0,0600	0,1100	0,1000	0,1700

Rótulos usados na Tabela 5: L: Linear; P: Polinomial; G: Gaussiano; PKKM: Procrustes kernel k-médias; PKKM_p: Procrustes kernel k-médias protótipo; VM: Voto majoritário; VP: Voto de pluralidade; VPP: Voto de pluralidade ponderado.

3.11 APLICAÇÃO EM DADOS REAIS

Os dados foram obtidos do pacote `shapes` no R e eles são descritos detalhadamente no livro de Dryden e Mardia (2016). A seguir, são descritos os dados utilizados nos métodos de classificação e todos eles serão ilustrados considerando as coordenadas procrustes completas (2.12):

1. **Crânios de Primatas - Gorila:** Os dados consistem em avaliar as diferenças cranianas entre os sexos de primatas. A amostra possui 29 gorilas macho e 30 gorilas fêmeas, todos na forma adulta, contendo 8 marcos. Na Figura 15 estão os marcos coletados dos crânios desses gorilas utilizando as coordenadas procrustes.

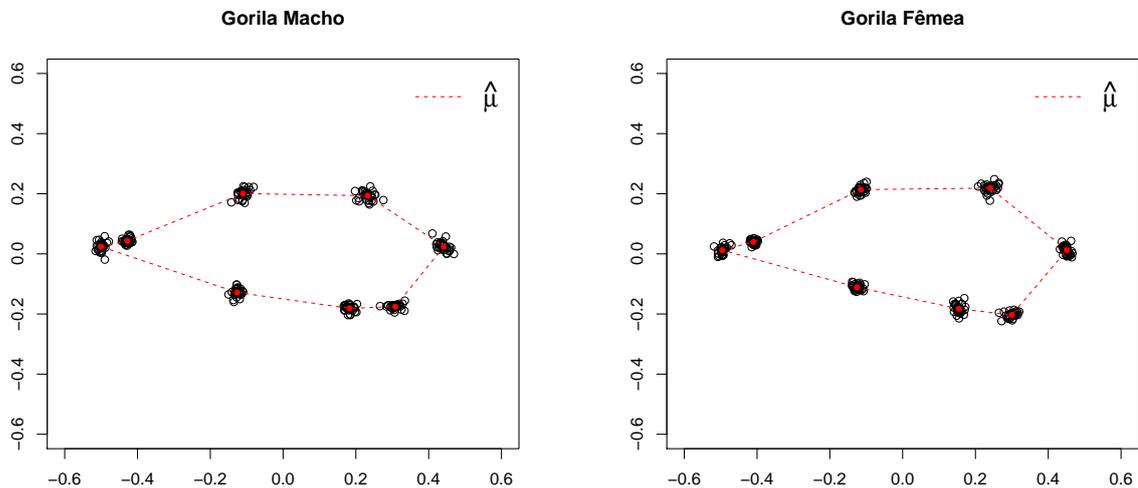


Figura 15 – Marcos nas coordenadas procrustes dos crânios de gorilas.

2. **Crânios de Primatas - Orangotango:** Como o mesmo intuito em avaliar diferenças cranianas entre os sexos de primatas, foram coletadas amostras de crânios de orangotangos adultos, que consistem em 24 fêmeas e 30 machos com 8 marcos. Na Figura 16 estão os marcos para esses dados utilizando as coordenadas procrustes.
3. **Vértebras de Ratos:** Esses dados consistem em um experimento para avaliar o peso corporal na forma das vértebras de ratos. Três grupos de ratos foram obtidos: grande, pequeno e controle. Os grupos grande e pequeno contém ratos selecionados de acordo com o peso corporal. Já o grupo controle, contém ratos não selecionados pelo peso. As vértebras são divididas em 30, 23 e 23 para os grupos controle, grande e pequeno, respectivamente. Ademais, cada vértebra possui 6 marcos. Consideramos as vértebras grandes e pequenas porque elas foram usadas anteriormente por Southworth, Mardia e

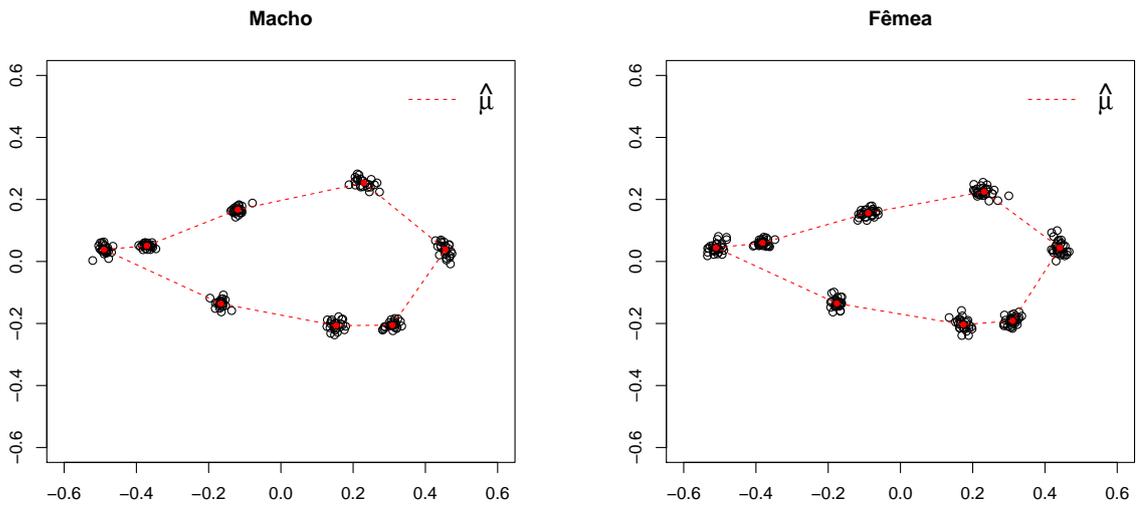


Figura 16 – Marcos nas coordenadas procrustes dos crânios de orangotangos.

Taylor (2000) e queremos comparar os resultados. Na Figura 17 estão os marcos para os dois grupos.

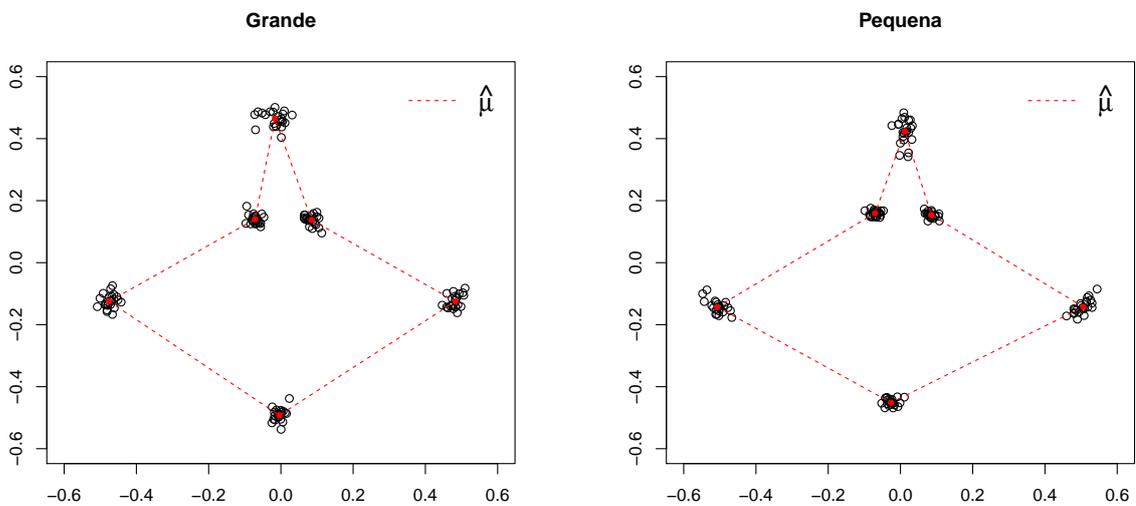


Figura 17 – Marcos nas coordenadas procrustes para as vértebras de ratos.

4. **Esquizofrenia:** Esse banco de dados possui 13 marcos tomados de ressonância magnética do cérebro de 14 voluntários do grupo controle e 14 pacientes com esquizofrenia. Esse experimento foi realizado com o intuito de averiguar diferenças entre as formas médias ou na variabilidade nessas duas situações. Na Figura 18 estão as informações sobre os marcos para ambos os grupos.

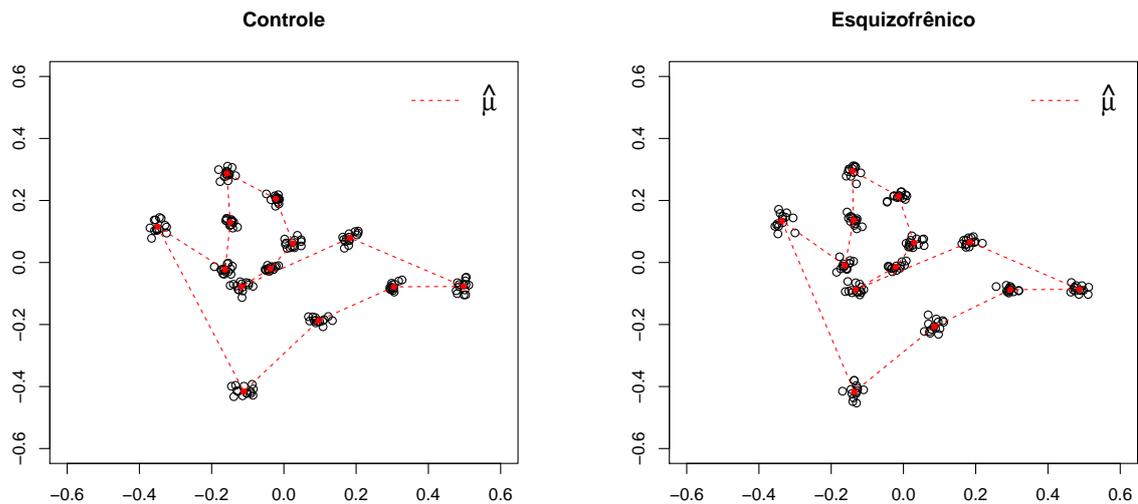


Figura 18 – Marcos nas coordenadas procrustes da ressonância magnética do cérebro de pacientes com e sem esquizofrenia.

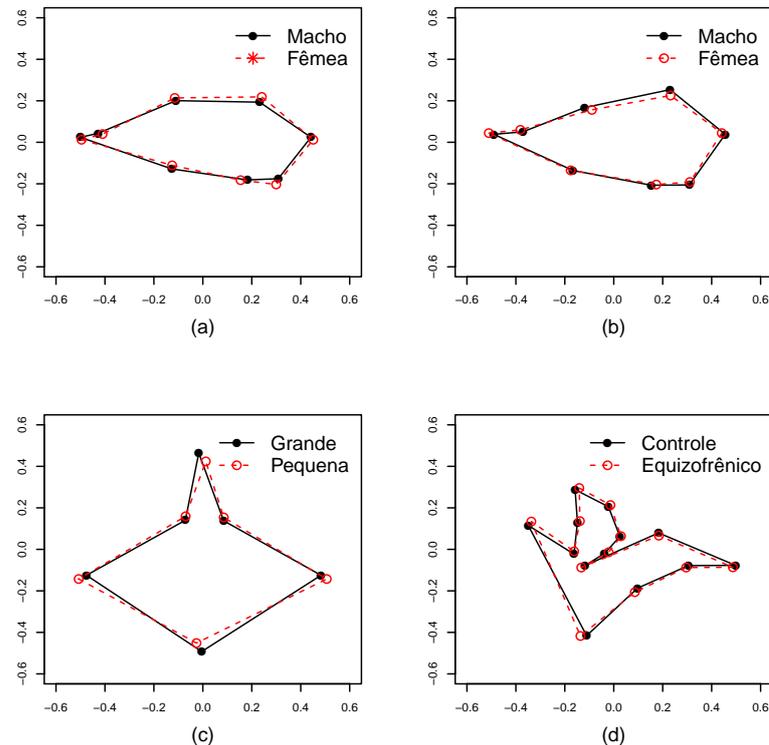


Figura 19 – Representação da forma média: (a) Crânio do gorila; (b) Crânio do orangotango; (c) Vértebra do rato; (d) Ressonância magnética do cérebro.

Na Figura 19 estão as formas médias dos grupos para cada conjunto de dados, em que pode-se verificar as diferenças entre as formas médias dos grupos. Na Tabela 6, estão as taxas de erros para todos os classificadores considerando as mesmas configurações utilizadas na simulação. A única diferença, é que foi utilizado o método *leave-one-out* devido ao fato de que os tamanhos de amostras são pequenos em cada conjunto de dados. Os métodos *ensemble* serão utilizados com os classificadores que obtiverem os melhores resultados.

- Gorila: O classificador HC através da estatística de James e Hotelling atingiram a menor taxa de erro que foi de 0,0%. Em segundo lugar, a SVM com e sem pesos e o HC com a estatística Lambda, com uma taxa de erro de 1,69%.

O classificador *ensemble* foi utilizado com os classificadores $SVM_C(d_F)$, $PKKM_p$, Bayes-ED com kernel normal, discriminante Bingham e o HC com a estatística de James. Em relação aos métodos de combinação do *ensemble*, todos produziram o mesmo resultado, com uma taxa de erro de 1,69%.

- Orangotango: A SVM com o kernel gaussiano produziu o melhor resultado, obtendo uma taxa de erro de 5,56%. O *ensemble* foi utilizado com os classificadores $SVM_C(d_F)$ com kernel gaussiano, $PKKM_p$, Bayes-ED com kernel t-Student, discriminante Bingham e o

HC com a estatística Lambda. Assim, o método VPP resultou no mesmo desempenho da SVM.

- Vértebra: Nesse caso a SVM com kernel polinomial alcançou uma taxa de erro de 0,0%. O *ensemble* foi utilizado com os classificadores $SVM_{\mathbb{C}}(d_F)$ com kernel polinomial, $PKKM_p$, Bayes-ED com kernel t-Student, discriminante Bingham e o HC com a estatística Lambda e o melhor resultado foi um erro de 4,35% para a VPP.
- Esquizofrenia: A SVM complexa também produziu o melhor resultado, atingiu uma taxa de erro de 14,29%. O *ensemble* foi utilizado com os classificadores $SVM_{\mathbb{C}}(d_F)$ com kernel polinomial, $PKKM_p$, Bayes-ED com kernel t-Student, discriminante Watson e o HC com a estatística de Goodall, o que resultou em uma taxa de erro de 17,86%.

Perceba que os classificadores não produziram resultados tão bons nos dados de esquizofrenia. O classificador de Bayes, utilizando a Bingham complexa, foi o que produziu o pior resultado, com uma taxa de erro de 75%. Em seguida, pelo classificador HC com as estatísticas de James e Hotelling, com taxas de erro de 50% e 53,35%, respectivamente. Um dos motivos é o tamanho amostral muito pequeno em cada grupo. Além disso, vários classificadores apresentaram superajuste (*overfitting*), ou seja, quando um classificar se ajusta muito bem ao conjunto de dados de treinamento, mas se mostra ineficaz para prever novos resultados. Entretanto, parece que a SVM complexa com o kernel polinomial conseguiu minimizar esse problema e alcançou a menor taxa de erro, que foi de 14,29%.

Southworth, Mardia e Taylor (2000) utilizaram os dados de vértebras para a classificação e a menor taxa de erro encontrada foi de 10,9%, utilizando uma rede neural invariante. Perceba que, no presente trabalho, todos os classificadores utilizados nos dados de vértebras, produziram uma taxa de erro menor.

Os classificadores também foram utilizados nos dados de vértebras utilizando os três grupos (grande, pequena e controle). Ademais, foi feita uma junção entre os dados de crânios de gorilas e orangotangos, formando um conjunto de dados, denominado de macacos, com 4 grupos. A Tabela 7 possui as taxas de erros utilizando esses dois conjuntos de dados. Veja que, para os dados de macacos, a menor taxa de erro foi para o classificador $PKKM_p$ com 0,88%, seguido pelo *ensemble* com um erro de 2,65%. Em relação aos dados de vértebras, a SVM e o Bayes-ED produziram a menor taxa de erro que foi de 1,32%.

Tabela 6 – Taxas de erro para cada classificador em relação aos 4 banco de dados, utilizando o método *leave-one-out*.

Classificador	Tipo	Gorila	Orangotango	Vértebra	Esquizofrenia
SVM _C	L	0,0169	0,0741	0,0435	0,2143
	P	0,0169	0,1111	0,0000	0,1429
	G	0,0169	0,0556	0,0217	0,2500
SVM _C (d_F)	L	0,0169	0,0926	0,0652	0,2143
	P	0,0169	0,0926	0,0000	0,1429
	G	0,0169	0,0556	0,0435	0,2500
SVM _C (d_P)	L	0,0169	0,0926	0,0652	0,2143
	P	0,0169	0,0926	0,0000	0,1429
	G	0,0169	0,0556	0,0435	0,2500
SVM _C (ρ)	L	0,0169	0,0926	0,0652	0,2143
	P	0,0169	0,0926	0,0000	0,1429
	G	0,0169	0,0556	0,0435	0,2500
Kernel k-médias	PKKM	0,0508	0,1481	0,0870	0,3929
	PKKM _p	0,0339	0,0926	0,0435	0,2500
Bayes-ED	Normal	0,0339	0,1852	0,0217	0,3929
	t-Student	0,0339	0,1296	0,0217	0,3929
Bayes	Bingham	0,0339	0,0926	0,0870	0,7500
	Watson	0,1017	0,1296	0,0870	0,3214
Hill-Climbing	James	0,0000	0,1111	0,0870	0,5000
	Hotelling	0,0000	0,1296	0,0870	0,5357
	Goodal	0,0847	0,1296	0,0870	0,3571
	Lambda	0,0169	0,0926	0,0652	0,4643
Ensemble	VM	0,0169	0,1667	0,0652	0,4286
	VP	0,0169	0,1667	0,0652	0,4286
	VPP	0,0169	0,0556	0,0435	0,1786

Rótulos usados na Tabela 6: L: Linear; P: Polinomial; G: Gaussiano; PKKM: Procrustes kernel k-médias; PKKM_p: Procrustes kernel k-médias protótipo; VM: Voto majoritário; VP: Voto de pluralidade; VPP: Voto de pluralidade ponderado.

3.12 CONCLUSÃO

Neste capítulo foram propostos vários métodos de classificação com uma ampla diversidade de técnicas para dados no espaço das pré-formas. Foi feito um estudo de simulação para avaliar os classificadores e esse estudo revelou o desempenho de cada classificador à medida que a variabilidade aumenta. Através desse estudo, pôde-se concluir que o desempenho diminui com grandes variabilidades e que a SVM complexa com e sem os pesos produziram resultados equivalentes. Além do mais, viu-se que a SVM com pesos é um pouco mais rápida e na grande maioria dos casos a SVM complexa com os kernels polinomial e gaussiano produziu os melhores resultados.

O classificador baseado no método *ensemble* reuniu os melhores classificadores e

Tabela 7 – Taxas de erro para cada classificador em relação aos dados de vértebras e macacos, utilizando o método *leave-one-out*.

Classificador	Tipo	Macacos	Vértebra
SVM _C	L	0,0531	0,0263
	P	0,0619	0,0658
	G	0,0442	0,0132
SVM _C (d_F)	L	0,0531	0,0395
	P	0,0619	0,0526
	G	0,0442	0,0263
SVM _C (d_P)	L	0,0531	0,0395
	P	0,0619	0,0526
	G	0,0531	0,0263
SVM _C (ρ)	L	0,0531	0,0395
	P	0,0619	0,0526
	G	0,0531	0,0263
Kernel k-médias	PKKM	0,0973	0,2368
	PKKM _p	0,0088	0,2237
Bayes-ED	Normal	0,1239	0,0132
	t-Student	0,1062	0,0132
Bayes	Bingham	0,0619	0,1974
	Watson	0,1239	0,2632
Hill-Climbing	Wilks	0,0442	0,2237
	Pillai	0,0796	0,2237
	Hotelling	0,0796	0,2368
	Roy	0,3274	0,3553
	Lambda	0,1858	0,2632
Ensemble	VM	0,0265	0,0921
	VP	0,0265	0,0921
	VPP	0,0265	0,0658

Rótulos usados na Tabela 6: L: Linear; P: Polinomial; G: Gaussiano; PKKM: Procrustes kernel k-médias; PKKM_p: Procrustes kernel k-médias protótipo; VM: Voto majoritário; VP: Voto de pluralidade; VPP: Voto de pluralidade ponderado.

não demonstrou um desempenho melhor do que alguns classificadores individuais. Dentre os métodos de combinação, o método do voto de pluralidade ponderado, foi o que produziu os melhores resultados.

As aplicações foram feitas em quatro bancos de dados e a SVM gerou os melhores resultados em três deles: crânio de orangotango, vértebra de rato e esquizofrenia. No qual, nos dados de vértebras a taxa de erro foi igual a zero. Já o classificador HC produziu uma taxa de erro igual a zero quando utilizado nos dados dos gorilas. Além do mais, os dados de esquizofrenia são os mais difíceis de classificar, devido ao baixo tamanho amostral e a própria natureza dos dados, e a SVM produziu a menor taxa de erro que foi de 14,29%, seguida pelo classificador *ensemble* com 17,86%. Além disso, os classificadores foram utilizados em problemas envolvendo três e

quatro grupos, em que os melhores resultados foram para a SVM, *ensemble* e $PKKM_p$.

4 BONDADDE DE AJUSTE PARA A DISTRIBUIÇÃO WATSON COMPLEXA

4.1 INTRODUÇÃO

A qualidade do ajuste de um modelo tem como objetivo descrever o quão bem o modelo se ajusta a um conjunto de dados. Essa qualidade do ajuste pode ser feita através de medidas, como os critérios AIC e BIC, ou ainda, através de testes estatísticos. Com o intuito de avaliar se determinada distribuição é adequada para um conjunto de dados, pode-se utilizar os seguintes testes: Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling e razão de verossimilhanças. Geralmente, as estatísticas desses testes possuem uma distribuição assintótica e podem não funcionar muito bem quando o tamanho da amostra é pequeno. Assim, pode-se utilizar o método *bootstrap* para gerar uma distribuição empírica e realizar o teste com base nessa distribuição.

As distribuições definidas em $\mathbb{C}S^{k-1}$ são de difícil manipulação, ocasionado em uma carência de testes e métodos de adequação. Pensando nisso, foi estudado o uso de uma aproximação da distribuição Watson complexa pela distribuição normal multivariada complexa singular e uma transformação que recai em uma distribuição Qui-quadrado que foi proposta por Mardia e Dryden (1999). Essa aproximação será utilizada para verificar adequação em dados para a distribuição Watson complexa.

A seguir será apresentada a distribuição Watson complexa com suas principais propriedades na Seção 4.2; na Seção 4.3 é mostrado um novo simulador para a distribuição Watson complexa; na Seção 4.4 é feita uma descrição sobre a distribuição normal multivariada singular e a sua versão complexa, além de algumas propriedades; na Seção 4.5 são definidos os testes de bondade de ajuste; na Seção 4.6 é mostrado uma aproximação da distribuição Watson complexa para a normal multivariada complexa singular e uma transformação que leva na distribuição de Dirichlet. Por fim, nas Seções 4.7 e 4.8 são mostrados resultados numéricos através de experimentos de simulação e aplicações em dados reais, respectivamente.

4.2 DISTRIBUIÇÃO WATSON COMPLEXA

A distribuição Watson complexa é uma importante e simples distribuição na esfera complexa para a análise de pré-formas. Evidentemente, essa distribuição é uma extensão da distribuição Watson real que é utilizada em dados direcionais. Além disso, a distribuição Watson complexa pode ser vista como um caso particular da distribuição Bingham complexa, que por

sua vez, é um caso especial da distribuição Bingham real (DRYDEN; MARDIA, 2016, p. 221).

A distribuição Watson complexa foi definida na subseção 3.2.3 no espaço das pré-formas $\mathbb{C}S^{k-2}$. Uma variável aleatória \mathbf{Z} possui distribuição Watson complexa, isto é, $\mathbb{C}W_{k-2}(\boldsymbol{\mu}, \kappa)$ se possuir a densidade

$$f(\mathbf{z}) = c(\kappa)^{-1} \exp \{ \kappa |\mathbf{z}^* \boldsymbol{\mu}|^2 \},$$

no qual $\boldsymbol{\mu}$ é a forma média, $\kappa \geq 0$ representa o parâmetro de concentração e $c(\kappa)$ é uma constante dada por

$$c(\kappa) = 2\pi^{k-1} \kappa^{2-k} \left\{ e^{\kappa} - \sum_{j=0}^{k-3} \frac{\kappa^j}{j!} \right\}.$$

Considerando uma amostra aleatória $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ da distribuição Watson complexa, temos que os estimadores de MV de $\boldsymbol{\mu}$ e κ são dados por

$$\hat{\boldsymbol{\mu}} = \mathbf{u}_1,$$

e considerando alta concentração, isto é, κ grande, tem-se

$$\hat{\kappa} \cong \frac{n(k-2)}{n - \lambda_1},$$

em que \mathbf{u}_1 é o autovetor correspondente ao maior autovalor λ_1 , ambos calculados através da decomposição espectral da matriz $\mathbf{S} = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^*$ (ver subseção 3.2.3).

4.3 SIMULADOR DA DISTRIBUIÇÃO WATSON COMPLEXA

Como mencionado, a distribuição Watson complexa é um caso particular da distribuição Bingham complexa. Assim, para simular dados da distribuição Watson complexa basta utilizar algum dos três geradores da Bingham que foram propostos por Kent, Constable e Er (2004). Entretanto, há uma limitação nesses geradores, pois a forma média é fixada em $\boldsymbol{\mu} = (0, \dots, 1)^\top$ (com $k-2$ zeros) e para que os dados possuam uma forma média diferente é necessário rotacioná-los.

O método da aceitação-rejeição funciona da seguinte maneira: dada uma densidade de interesse $f(x)$, é preciso encontrar uma densidade $g(x)$ e uma constante M tal que

$$f(x) \leq Mg(x), \quad (4.1)$$

no qual $g(x)$ possui o mesmo (ou maior) suporte de $f(x)$ e $M \geq 1$. Assim, o algoritmo de geração de números aleatórios pelo método da aceitação-rejeição é dado por

1. Gere de forma independente $U \sim \mathcal{U}(0, 1)$ (distribuição uniforme no intervalo $(0, 1)$) e $X \sim g$;
2. Aceite $V = X$ (aceite X como ocorrência de f) se

$$U \leq \frac{f(X)}{Mg(X)};$$

3. Caso contrário, retorne ao passo 1.

Dessa forma, é proposto um gerador utilizando o método da aceitação-rejeição para a distribuição Watson complexa, no qual o usuário poderá utilizar a forma média desejada. Portanto, não sendo necessário rotacionar os dados, facilitando o processo e controle da simulação. Com esse intuito, será utilizado uma desigualdade proposta por Kent, Ganeiber e Mardia (2013), que é dada por

$$e^{-u} \leq e^{-(q-b)/2} \left(\frac{q/b}{1 + 2u/b} \right)^{q/2}, \quad (4.2)$$

com $u > 0$, $q > 0$ e $0 < b < q$. No qual, (4.2) deve ser minimizada em relação a b . Essa desigualdade foi utilizada para gerar amostras da Bingham real através da distribuição Gaussiana angular central (GAC), conseqüentemente, também pode ser utilizada para gerar dados da Bingham complexa e várias outras distribuições relacionadas.

A distribuição GAC é uma alternativa para a distribuição Bingham para dados direcionais. Em Tyler (1987) foram estudadas algumas propriedades, como estimação via MV, teste de uniformidade, teste de circularidade e análise de componentes principais. Uma variável aleatória $\mathbf{X} \in S^{p-1} = \{\mathbf{X} : \|\mathbf{X}\| = 1, \mathbf{X} \in \mathbb{R}^p\}$ com distribuição GAC(\mathbf{A}), possui a densidade

$$f(\mathbf{x}) = \frac{\Gamma(\frac{p}{2})}{2\pi^{p/2}} |\mathbf{A}|^{-1/2} \left(\mathbf{x}^\top \mathbf{A}^{-1} \mathbf{x} \right)^{-p/2},$$

em que \mathbf{A} é uma matriz simétrica e positiva definida. Existe também a sua versão para a esfera complexa. Considere $\mathbf{Z} \in \mathbb{C}S^{p-1} = \{\mathbf{Z} : \|\mathbf{Z}\| = 1, \mathbf{Z} \in \mathbb{C}^p\}$, então a distribuição complexa gaussiana angular central (CGAC) é dada por

$$f(\mathbf{z}) = \frac{\Gamma(p)}{2\pi^p} |\mathbf{A}|^{-1} (\mathbf{z}^* \mathbf{A}^{-1} \mathbf{z})^{-p}, \quad (4.3)$$

em que \mathbf{A} é uma matriz Hermitiana positiva definida (DRYDEN; MARDIA, 2016, p. 231). Para utilizar essa distribuição para os dados das pré-formas, basta fazer $p = k - 1$, Note que a distribuição CGAC pode ser vista como uma distribuição GAC de dimensão $2p$. Outro fato importante, é que para gerar uma amostra aleatória da distribuição GAC(\mathbf{A}), basta gerar $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{A})$ e utilizar a transformação $\mathbf{X}/\|\mathbf{X}\| \sim \text{GAC}(\mathbf{A})$ (MARDIA; JUPP, 2000, p. 182).

Considere uma variável aleatória $\mathbf{Z} \sim \mathbb{C}W_{k-2}(\boldsymbol{\mu}, \kappa)$. Assim, a densidade é dada por

$$f(\mathbf{z}) = c(\kappa)^{-1} e^{\kappa|\mathbf{z}^* \boldsymbol{\mu}|^2},$$

em que

$$c(\kappa) = 2\pi^{k-1} \kappa^{2-k} \left(e^\kappa - \sum_{j=0}^{k-3} \frac{\kappa^j}{j!} \right).$$

É possível reescrever a constante $c(\kappa)$ da seguinte forma

$$\begin{aligned} c(\kappa) &= 2\pi^{k-1} \kappa^{2-k} \left(e^\kappa - \sum_{j=0}^{k-3} \frac{\kappa^j}{j!} \right), \\ &= 2\pi^{k-1} \kappa^{2-k} e^\kappa \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right). \end{aligned}$$

Assim,

$$\begin{aligned} f(\mathbf{z}) &= \frac{e^{\kappa|\mathbf{z}^* \boldsymbol{\mu}|^2}}{2\pi^{k-1} \kappa^{2-k} e^\kappa \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right)} = \frac{e^{\kappa|\mathbf{z}^* \boldsymbol{\mu}|^2 - \kappa}}{2\pi^{k-1} \kappa^{2-k} \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right)}, \\ &= \frac{e^{-\kappa(1-|\mathbf{z}^* \boldsymbol{\mu}|^2)}}{2\pi^{k-1} \kappa^{2-k} \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right)} = c_1(\kappa)^{-1} e^{-\kappa(1-|\mathbf{z}^* \boldsymbol{\mu}|^2)}, \end{aligned}$$

em que $c_1(\kappa) = 2\pi^{k-1} \kappa^{2-k} \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right)$. Como, $|\mathbf{z}^* \boldsymbol{\mu}|^2 = \mathbf{z}^* \boldsymbol{\mu} \boldsymbol{\mu}^* \mathbf{z}$, então a densidade pode ser reescrita como

$$\begin{aligned} f(\mathbf{z}) &= c_1(\kappa)^{-1} e^{-\kappa(1-\mathbf{z}^* \boldsymbol{\mu} \boldsymbol{\mu}^* \mathbf{z})}, \\ &= c_1(\kappa)^{-1} e^{-\kappa(\mathbf{z}^* \mathbf{z} - \mathbf{z}^* \boldsymbol{\mu} \boldsymbol{\mu}^* \mathbf{z})}, \quad \text{pois } \mathbf{z}^* \mathbf{z} = 1 \\ &= c_1(\kappa)^{-1} e^{-\kappa \mathbf{z}^* (\mathbf{I}_{k-1} - \boldsymbol{\mu} \boldsymbol{\mu}^*) \mathbf{z}}. \end{aligned}$$

Nesse caso, tome $u = \kappa \mathbf{z}^* (\mathbf{I}_{k-1} - \boldsymbol{\mu} \boldsymbol{\mu}^*) \mathbf{z}$. Então, pela desigualdade (4.2), temos

$$\begin{aligned} e^{-u} &\leq e^{-(q-b)/2} \left(\frac{q/b}{1+2u/b} \right)^{q/2}, \\ &\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} \left(\frac{1}{\mathbf{z}^* [\mathbf{I}_{k-1} + 2\kappa(\mathbf{I}_{k-1} - \boldsymbol{\mu} \boldsymbol{\mu}^*)/b] \mathbf{z}} \right)^{q/2}, \\ &\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} \left(\frac{1}{\mathbf{z}^* \mathbf{A}^{-1} \mathbf{z}} \right)^{q/2}, \quad \mathbf{A}^{-1} = \mathbf{I}_{k-1} + 2\kappa(\mathbf{I}_{k-1} - \boldsymbol{\mu} \boldsymbol{\mu}^*)/b. \end{aligned}$$

Portanto,

$$\begin{aligned} \frac{c_1(\boldsymbol{\kappa})}{c_1(\boldsymbol{\kappa})} e^{-\kappa \mathbf{z}^* (\mathbf{I}_{k-1} - \boldsymbol{\mu} \boldsymbol{\mu}^*) \mathbf{z}} &\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} (\mathbf{z}^* \mathbf{A}^{-1} \mathbf{z})^{-q/2} \frac{|\mathbf{A}|^{-1} \Gamma(q/2)}{|\mathbf{A}|^{-1} 2\pi^{q/2}} \frac{2\pi^{q/2}}{\Gamma(q/2)}, \\ f(\mathbf{z}) &\leq c_1(\boldsymbol{\kappa})^{-1} e^{-(q-b)/2} q^{q/2} b^{-q/2} \frac{2\pi^{q/2}}{\Gamma(q/2)} |\mathbf{A}| g(\mathbf{z}), \end{aligned}$$

em que $f(\mathbf{z})$ é a densidade de uma variável com distribuição Watson complexa e $g(\mathbf{z})$ com densidade de uma variável com distribuição CGAC(\mathbf{A}) fazendo $q = 2(k-1)$. Desse modo, foi encontrada uma desigualdade no formato de (4.1), no qual

$$M = c_1(\boldsymbol{\kappa})^{-1} e^{-(q-b)/2} q^{q/2} b^{-q/2} \frac{2\pi^{q/2}}{\Gamma(q/2)} |\mathbf{A}|, \quad (4.4)$$

que é uma função não-linear em função de um único parâmetro desconhecido (b), que deve ser minimizada utilizando métodos numéricos. Ademais, $1/M$ representa a probabilidade de aceitação do algoritmo. O Algoritmo 14 resume os passos para a geração de n amostras da distribuição Watson complexa pelo método da aceitação-rejeição.

Na Tabela 8 estão as probabilidades de aceitação para vários valores da dimensão do vetor de pré-forma ($k-1$), com a forma média fixada em $\boldsymbol{\mu} = (0, \dots, i)^\top$ (o número de zeros é igual a $k-2$) e diferentes valores de κ utilizando o Algoritmo 14. Pode-se ver que há uma probabilidade de aceitação elevada (acima de 70%) para os valores de $\kappa = 1, 3, 5$ e 8, em relação a todos os valores da dimensão do vetor. Veja ainda que, comparando o aumento de κ para cada valor de k , pode-se ver um decaimento linear, ou seja, à medida que o valor de κ aumenta, a probabilidade de aceitação vai diminuindo para todos os valores de k . Esse comportamento ocorre, devido ao fato de que quanto maior o valor de κ , maior a concentração das observações em torno da forma média, assim, as observações precisam ser mais homogêneas para serem aceitas. O contrário ocorre quando os valores de κ são pequenos, ou seja, valores pequenos de κ indicam grandes variabilidades e, conseqüentemente, uma alta probabilidade de aceitação.

Tabela 8 – Probabilidade de aceitação ($1/M$) utilizando a distribuição CGAC para gerar amostras para a distribuição Watson complexa, utilizando vários valores do parâmetro κ e da dimensão do espaço de configuração k , para o vetor de médias $\boldsymbol{\mu} = (0, \dots, i)^\top$ com $k - 2$ zeros.

κ	k							
	3	4	5	6	7	8	9	10
1	0,9808	0,9895	0,9944	0,9968	0,9980	0,9987	0,9991	0,9994
3	0,8914	0,9084	0,9385	0,9609	0,9750	0,9835	0,9888	0,9921
5	0,8223	0,8081	0,8384	0,8782	0,9133	0,9395	0,9577	0,9700
8	0,7693	0,7103	0,7076	0,7327	0,7717	0,8150	0,8556	0,8901
10	0,7508	0,6743	0,6530	0,6605	0,6861	0,7229	0,7651	0,8075
20	0,7144	0,6064	0,5511	0,5203	0,5037	0,4967	0,4972	0,5038
30	0,7026	0,5857	0,5220	0,4824	0,4563	0,4387	0,4270	0,4197
40	0,6968	0,5758	0,5084	0,4652	0,4354	0,4139	0,3981	0,3863
50	0,6933	0,5699	0,5005	0,4554	0,4236	0,4002	0,3824	0,3685
100	0,6864	0,5584	0,4852	0,4367	0,4017	0,3751	0,3541	0,3371
100	0,6803	0,5484	0,4722	0,4210	0,3836	0,3548	0,3318	0,3127
5000	0,6797	0,5476	0,4710	0,4196	0,3821	0,3531	0,3299	0,3107

Algoritmo 14: Simulador para a distribuição Watson complexa utilizando o método da aceitação-rejeição.

Entrada: Número de amostras n , valor de κ e a forma média $\boldsymbol{\mu}$.

Saída: n amostras da distribuição Watson complexa.

Sejam $f(\mathbf{z})$ e $g(\mathbf{z})$ as densidades da distribuição Watson complexa e CGAC, respectivamente.

Gere de forma independente uma amostra u de $U \sim \mathcal{U}(0, 1)$ e \mathbf{z} de $\mathbf{Z} \sim g$.

Aceite \mathbf{z} como sendo uma amostra da Watson complexa se

$$u \leq \frac{f(\mathbf{z})}{Mg(\mathbf{z})},$$

em que M é dado por (4.4) e deve ser minimizado em relação a b . Caso contrário, volte ao passo 2.

Repita os passos 2 e 3 até que se tenha n amostras.

Retorne as n amostras e a probabilidade de aceitação $1/M$.

Para um melhor entendimento desse processo de geração de amostras no espaço das pré-formas, vamos considerar a seguinte matriz de configuração

$$\tilde{\mathbf{Y}}_{\mu} = \begin{pmatrix} 4 & 3 \\ 8 & 3 \\ 6 & 7 \end{pmatrix}. \quad (4.5)$$

Na Figura 20 está a forma geométrica gerada por essas coordenadas. Perceba que a forma geométrica possui 3 marcos e representa um triângulo. O próximo passo é transformar a matriz $\tilde{\mathbf{Y}}_{\mu}$ (4.5) em um vetor complexo, remover os efeitos de locação e escala (para mais detalhes ver Capítulo 2). Preliminarmente, vamos transformar a matriz $\tilde{\mathbf{Y}}_{\mu}$ no vetor complexo

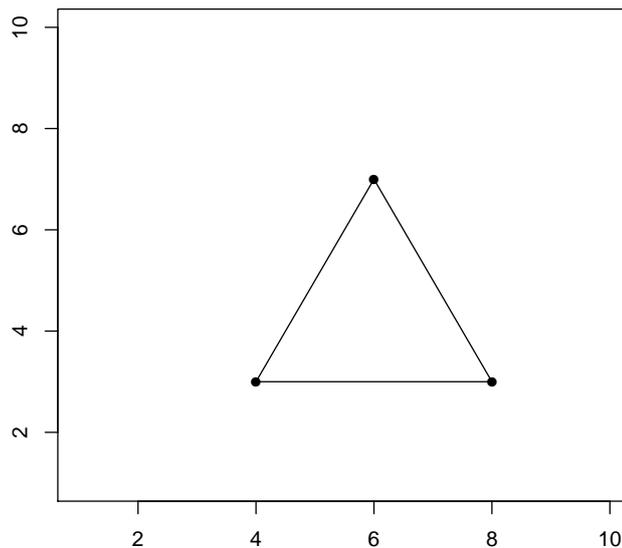


Figura 20 – Forma média com 3 marcos.

$$\mathbf{z}_{\mu} = (4 + 3i, 8 + 3i, 6 + 7i)^{\top}. \quad (4.6)$$

Para remover o efeito da locação é necessário pré-multiplicar (4.6) pela submatriz de Helmert

$$\boldsymbol{\omega}_{\mu} = \mathbf{H}\mathbf{z}_{\mu}. \quad (4.7)$$

Veja que nessa operação uma dimensão é perdida. Dessa forma, pré-multiplicaremos (4.7) novamente por \mathbf{H}^{\top} . Assim, temos que

$$\boldsymbol{\omega}_{\mu_2} = \mathbf{H}^\top \mathbf{H} \mathbf{z}_\mu. \quad (4.8)$$

Finalmente, para remover o efeito de escala, basta dividir (4.8) pela sua norma. Portanto, obtendo a forma média no espaço das pré-formas que é dada por

$$\boldsymbol{\mu} = \frac{\boldsymbol{\omega}_{\mu_2}}{\|\boldsymbol{\omega}_{\mu_2}\|} = \begin{pmatrix} -0,46291 - 0,3086067i \\ 0,46291 - 0,3086067i \\ 0,00000 + 0,6172134i \end{pmatrix}. \quad (4.9)$$

Para manter a mesma notação utilizada anteriormente para as pré-formas, considere que nesse caso $k - 1 = 3$, pois estamos simulando dados diretamente do espaço das pré-formas. Na Figura 21 está representado o triângulo (4.9) no espaço das pré-formas. Desta maneira, pode-se utilizar a forma média descrita em (4.9), fornecer κ e n no Algoritmo 14 e gerar amostras da distribuição Watson complexa. Para criar qualquer outra forma média, basta repetir esses procedimentos. Pode-se ver na Figura 22 alguns exemplos da utilização do simulador com um tamanho amostral $n = 50$, utilizando a forma média da equação (4.9) e diferentes valores de κ . Perceba que à medida que o valor de κ aumenta, os pontos ficam mais concentrados em torno dos marcos da forma média.

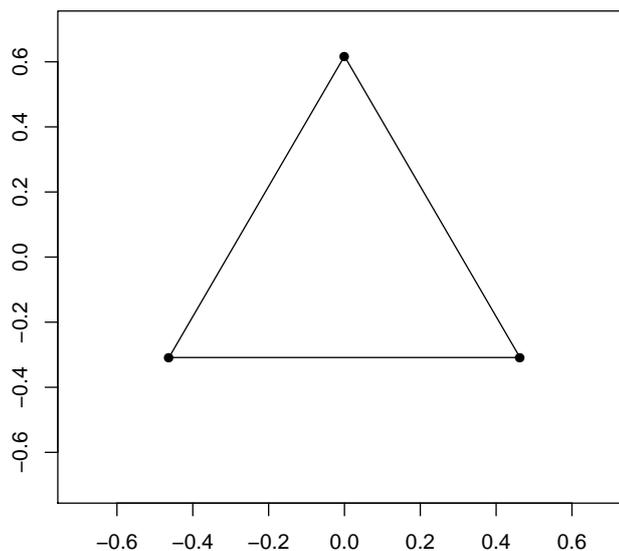


Figura 21 – Forma média com 3 marcos no espaço das pré-formas.

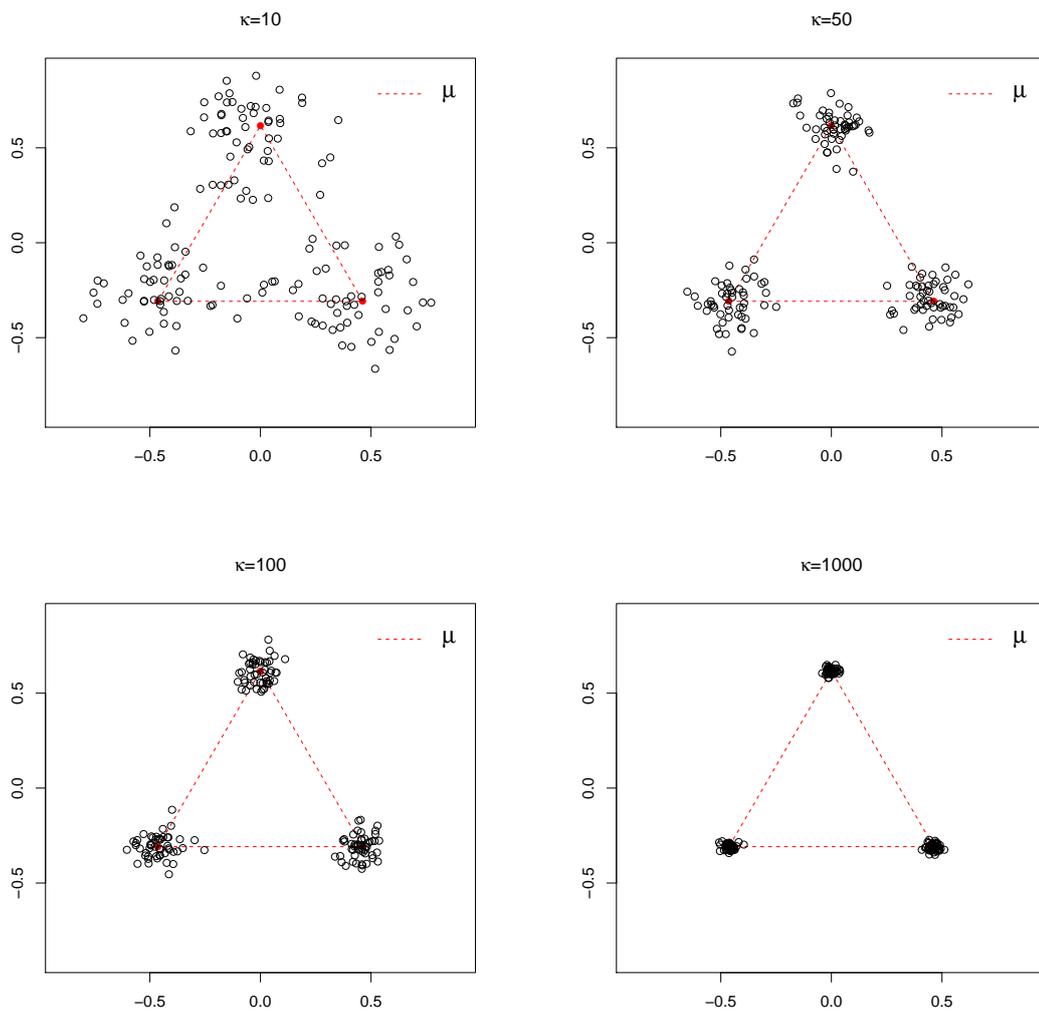


Figura 22 – Simulação no espaço das pré-formas com $n = 50$, forma média μ da equação (4.9) e utilizando $\kappa = 10, 50, 100, 1000$.

4.4 DISTRIBUIÇÃO NORMAL MULTIVARIADA

A distribuição normal pode ser considerada como a distribuição mais utilizada e a mais importante da Estatística. Essa distribuição também é conhecida como distribuição Gaussiana ou de Gauss, em homenagem ao matemático Karl Friedrich Gauss. Entretanto, Abraham de Moivre foi quem descobriu a distribuição normal (HALD, 2003, p. 469). A distribuição normal multivariada é uma extensão natural do caso univariado. A sua utilização também se deve as várias propriedades, dadas pelo fato de que é uma distribuição matematicamente tratável, pois produz uma infinidade de resultados importantes, diferentemente de outras distribuições multivariadas. Evidentemente, ter boas propriedades matemáticas por si só não basta. Johnson e Wichern (2007, p. 149) citam razões para o uso da distribuição normal, 1º: a distribuição normal pode servir como um modelo em alguns casos; 2º: a distribuição amostral de muitas estatísticas multivariadas são aproximadamente normais, independentemente da forma da população parental, devido ao efeito do Teorema Central do Limite.

Seja $\mathbf{X} = (X_1, \dots, X_p)^\top$ um vetor aleatório tendo distribuição normal multivariada com $\boldsymbol{\mu} = E(\mathbf{X})$ representando o vetor de médias e $\boldsymbol{\Sigma} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top]$ a matriz de covariância, sendo uma matriz positiva definida. Então, a densidade de \mathbf{X} é dada por

$$f(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}}.$$

Seja \mathbf{X} um vetor aleatório com distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Assim, são válidas as seguintes propriedades:

1. Se \mathbf{a} é um vetor de constantes, então $\mathbf{a}^\top \mathbf{X} = \sum_{i=1}^p a_i X_i$ tem distribuição $N(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a})$;
2. Se $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$, isso implica que cada X_1, \dots, X_p são independentes;
3. Se $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\mathbf{Y} = (\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim N_p(0, \mathbf{I})$, no qual, $\boldsymbol{\Sigma}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top$ utilizando a decomposição espectral, em que $\boldsymbol{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$. Essa operação é conhecida como raiz quadrada de uma matriz;
4. A distribuição Qui-quadrado com p graus de liberdade (χ_p^2) pode ser obtida por meio da soma do quadrado de variáveis independentes normais padrão, ou seja, $\sum_{i=1}^p Y_i^2$, em que $Y_i \sim N(0, 1)$ ($i = 1, 2, \dots, p$). Assim,

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2; \quad (4.10)$$

5. Se $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{I})$ e \mathbf{B} é uma matriz de posto r idempotente, ou seja, $\mathbf{B}\mathbf{B} = \mathbf{B}$, então

$$(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{B}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_r^2.$$

As demonstrações dessas e outras propriedades podem ser vistas em Rencher e Schaalje (2008).

4.4.1 Distribuição Normal Multivariada Singular

Em algumas situações a matriz Σ pode ser singular. Assim, é necessário uma mudança na densidade da normal multivariada. Seja $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ e o posto de Σ é igual a $s < p$. A densidade da distribuição normal multivariada singular é definida por

$$f(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-} (\mathbf{x} - \boldsymbol{\mu})\}}{(2\pi)^{s/2} (\prod_{i=1}^s \lambda_i)^{1/2}}, \quad (4.11)$$

no qual $\lambda_1, \dots, \lambda_s$ são os autovalores maiores que zero de Σ e Σ^{-} é a inversa generalizada de Moore-Penrose (RAO, 1973, p. 527-528). O vetor \mathbf{x} encontra-se no hiperplano $\mathbf{N}^\top (\mathbf{x} - \boldsymbol{\mu}) = 0$, no qual \mathbf{N} é uma matriz $p \times (p - s)$ tal que

$$\mathbf{N}^\top \Sigma = \mathbf{0} \quad \text{e} \quad \mathbf{N}^\top \mathbf{N} = \mathbf{I}_{p-s}. \quad (4.12)$$

Existe uma conexão entre as densidades da normal multivariada singular e da não-singular, descrita no Teorema a seguir.

Teorema 4. Considere $\mathbf{Y} \sim N_s(\mathbf{0}, \Lambda_1)$, em que Λ_1 é uma matriz diagonal de dimensão $s \times s$ cujo os elementos são $\lambda_i > 0$ com $i = 1, 2, \dots, s$. Então, existe uma matriz $(p \times s)$ \mathbf{C} de colunas ortogonais (isto é, $\mathbf{C}\mathbf{C}^\top = \mathbf{I}_s$) tal que

$$\mathbf{X} = \mathbf{C}\mathbf{Y} + \boldsymbol{\mu}, \quad (4.13)$$

tem densidade (4.11).

Para provar esse resultado é necessário uma generalização do método Jacobiano aplicado em hipersuperfícies (MARDIA; KENT; BIBBY, 1979, p. 42).

Como ilustração, considere uma amostra de $n = 300$ de variável aleatória $\mathbf{Y} \sim N_2(\mathbf{0}, \Sigma)$ com distribuição normal bivariada, com $\Sigma = \mathbf{I}$. Além disso, seja $\boldsymbol{\mu} = (2, 2, 2)^\top$ e considere a matriz

$$C = \begin{pmatrix} -0,5199416 & 0,6295454 \\ -0,8051731 & -0,1355099 \\ -0,2852315 & -0,7650553 \end{pmatrix}.$$

Utilizando a transformação para as amostras geradas através da equação (4.13), temos na Figura 23 a representação dos dados simulados no plano tridimensional. Observe que os dados se encontram em cima de um plano.

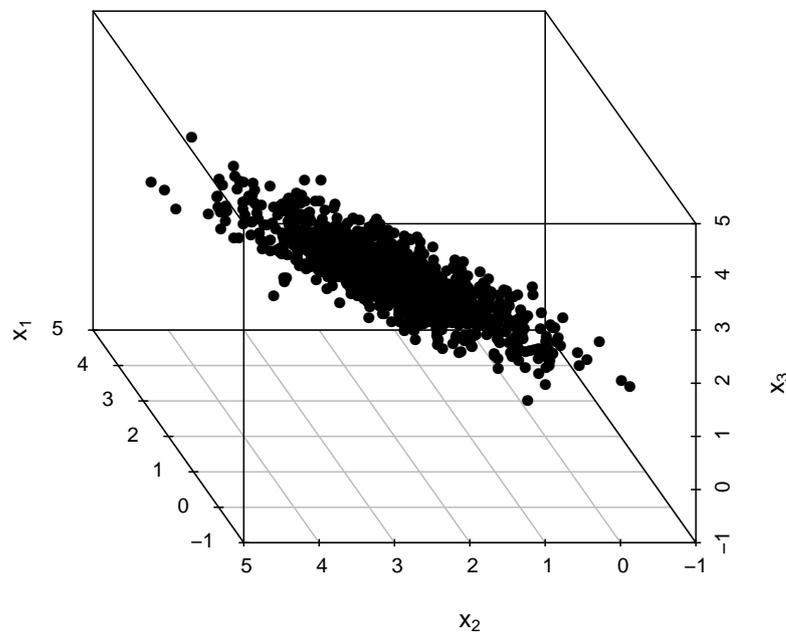


Figura 23 – Representação das amostras de uma distribuição normal multivariada singular.

Usando o resultado do Teorema 4, pode-se mostrar diversas propriedades da distribuição normal multivariada singular, dentre elas, pode-se mostrar que

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_s^2. \quad (4.14)$$

Considere agora a transformação da equação (4.14) para os dados da Figura 23. Assim, obtém-se uma distribuição Qui-quadrado com 2 graus de liberdade, como mostrado na Figura 24. Para mais detalhes sobre a distribuição normal multivariada singular ver Khatri (1968) e Tsukuma e Kubokawa (2015).

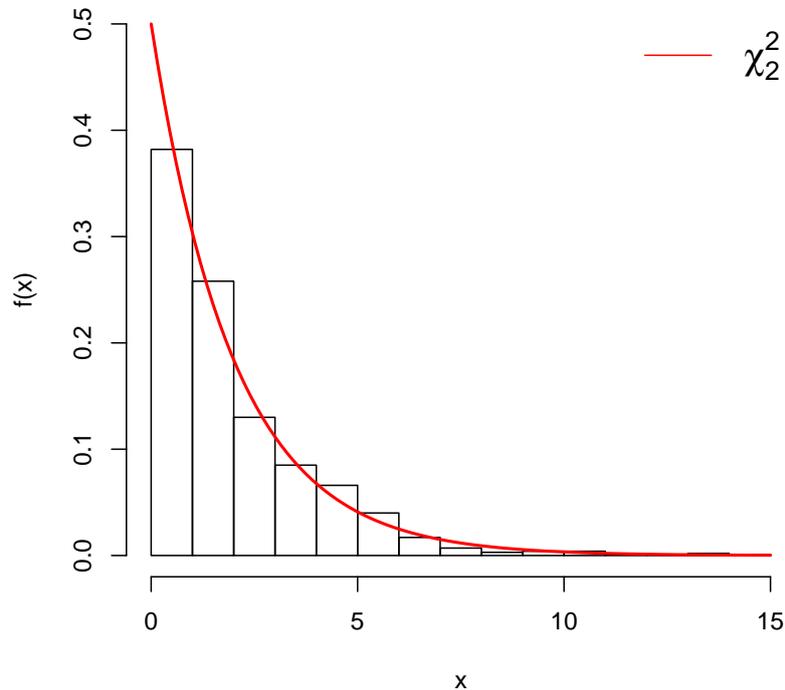


Figura 24 – Histograma das amostras de uma distribuição normal multivariada singular utilizando a transformação (4.14)

4.4.2 Distribuição Normal Multivariada Complexa

Uma variável aleatória normal complexa $Z = X + iV$, é uma variável aleatória em que as partes real e imaginária possuem uma distribuição normal bivariada. Assim, uma variável aleatória complexa multivariada p -dimensional, nada mais é que, uma normal multivariada $2p$ -dimensional (GOODMAN, 1963). Assim, pode-se concluir que a distribuição normal complexa é uma extensão da distribuição normal para os números complexos.

Seja $\mathbf{Z} = \mathbf{X} + i\mathbf{V}$ um vetor complexo p -dimensional tendo distribuição normal multivariada complexa, isto é, $\mathbf{Z} \sim \mathbb{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então a densidade de \mathbf{Z} é dada por

$$f(\mathbf{z}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^* \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})\right\}}{(2\pi)^p |\boldsymbol{\Sigma}|},$$

em que $\boldsymbol{\mu} = \boldsymbol{\mu}_x + i\boldsymbol{\mu}_v$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{vv} + i(\boldsymbol{\Sigma}_{vx} - \boldsymbol{\Sigma}_{xv})$ e $\boldsymbol{\Sigma}_{xv} = -\boldsymbol{\Sigma}_{vx}$ é uma matriz anti-simétrica. Para mais detalhes ver Goodman (1963) e Wooding (1956).

Como existe uma relação entre a distribuição normal multivariada e a normal multivariada complexa, pode-se utilizar uma matriz de covariância singular. Desse modo, considere

uma variável aleatória \mathbf{Z} com distribuição normal multivariada complexa singular, diga-se $\mathbf{Z} \sim \mathbb{CSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com a matriz de covariância $\boldsymbol{\Sigma}$ de posto s ($s < p$), então a densidade é dada por

$$f(\mathbf{z}) = \frac{\exp\{-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^* \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})\}}{(2\pi)^s (\prod_{i=1}^s \lambda_i)}.$$

Consequentemente, pode-se mostrar que

$$(\mathbf{Z} - \boldsymbol{\mu})^* \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}) \sim \chi_{2s}^2. \quad (4.15)$$

4.5 TESTES DE HIPÓTESES

Em muitas situações é de interesse do pesquisador realizar testes sobre os parâmetros da população com base nas informações obtidas na amostra. Assim sendo, surge na Estatística, métodos de inferência com o intuito de testar hipóteses. Os testes de hipóteses ajudam o pesquisador a tomar uma decisão, em que ele terá que optar em rejeitar ou não determinada afirmação baseando-se em um conjunto de evidências.

4.5.1 Teste da Razão de Verossimilhanças

O teste da razão de verossimilhanças (TRV) para realizar um teste de hipóteses está diretamente relacionado com os estimadores de MV, discutido na subseção 3.2.1. Sejam X_1, \dots, X_n uma amostra aleatória de uma população com densidade $f(x; \theta)$. Seja θ um parâmetro (ou vetor) de uma população e que deseja-se testar, $\mathbf{H}_0 : \theta \in \Theta_0$ e $\mathbf{H}_1 : \theta \in \Theta_1$, no qual $\Theta = \Theta_0 \cup \Theta_1$ e $\Theta_0 \cap \Theta_1 = \emptyset$. A estatística do TRV é definida como (ver Casella e Berger (2002, p. 375))

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})}, \quad (4.16)$$

em que a região crítica é $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ com $0 \leq c \leq 1$. Obviamente, o supremo de cada verossimilhança em (4.16) será justamente o EMV restrito a cada espaço, ou seja,

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})},$$

no qual $\hat{\theta}_0$ é o EMV de θ restrito a Θ_0 e $\hat{\theta}$ é o EMV irrestrito. Se \mathbf{H}_0 é verdadeira, espera-se que (4.16) seja próximo de 1, caso contrário, se (4.16) é próximo de 0, tem-se fortes evidências que \mathbf{H}_0 é falsa, por conseguinte, deve ser rejeitada. Para determinar o valor de c é preciso resolver a equação

$$\alpha = \sup_{\Theta_0} P(\lambda(\mathbf{X}) \leq c).$$

Para resolver esse problema, é necessário conhecer a distribuição de $\lambda(\mathbf{X})$, que na maioria das vezes, não é um problema fácil. Entretanto, pode-se encontrar uma função s estritamente crescente no domínio de $\lambda(\mathbf{X})$, de tal maneira que, $s(\lambda(\mathbf{X}))$ tenha uma forma simples com uma distribuição conhecida sob \mathbf{H}_0 (BOLFARINE; SANDOVAL, 2010, p. 136). No entanto, muitas vezes isso pode não ser trivial. Dessa forma, sob condições bem gerais pode-se utilizar a estatística

$$-2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi_{v,\alpha}^2 \quad (4.17)$$

em que v é o número de parâmetros que estão sendo testados em \mathbf{H}_0 (CASELLA; BERGER, 2002, p. 489). A hipótese nula é rejeitada se $-2 \log \lambda(\mathbf{x}) \geq \chi_{v,\alpha}^2$.

4.5.2 Teste de Kolmogorov

O teste para uma amostra de Kolmogorov é utilizado para chegar a aderência de distribuições, ou seja, averiguar se um conjunto de dados possui uma distribuição específica. O teste verifica o grau de concordância entre a distribuição empírica e uma distribuição contínua teórica estabelecida. As hipóteses do teste de Kolmogorov são

$$\begin{cases} \mathbf{H}_0 : \text{os dados possuem distribuição } F; \\ \mathbf{H}_1 : \text{os dados não possuem distribuição } F. \end{cases} \quad (4.18)$$

Para testar essas hipóteses, o teste se baseia nas diferenças entre os valores da função de distribuição empírica e teórica. Sejam X_1, \dots, X_n uma amostra aleatória de certa população, com função de distribuição $F_0(x)$. Assim, tomemos $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ como sendo as observações aleatórias ordenadas de forma crescente. A estatística do teste de Kolmogorov é dada por

$$D_n = \max_{1 \leq i \leq n} \left\{ F_0(x_{(i)}) - \frac{i}{n}, \frac{i-1}{n} - F_0(x_{(i)}) \right\}. \quad (4.19)$$

Observe que D_n representa a maior distância entre as funções de distribuição teórica e empírica, dando uma ideia do ajustamento. Sob \mathbf{H}_0 , Kolmogorov (1933) mostrou que a função de distribuição assintótica da estatística (4.19) é dada por

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n \leq x) = 1 + 2 \sum_{j=1}^{\infty} (-1)^j e^{-2j^2 x^2}.$$

Smirnov (1939) estendeu o teste de Kolmogorov para duas amostras independentes. Além disso, também para testes unilaterais, isto é, testar se uma das populações possui uma função de distribuição maior ou menor do que a outra. Frequentemente, esses dois testes são referidos como o teste de Kolmogorov-Smirnov (TKS) para uma ou duas amostras independentes.

4.5.3 Teste de Cramér-von Mises

O teste para uma amostra de Cramér-von Mises (TCM) também é utilizado para testar adequação de distribuições e as hipóteses do teste de são as mesmas do TKS (ver (4.18)). Sejam X_1, \dots, X_n uma amostra aleatória, considere ainda a amostra ordenada $X_{(1)}, \dots, X_{(n)}$, então a estatística de teste é dada por

$$C_n = \frac{1}{12n} + \sum_{i=1}^n \left(F(x_{(i)}) - \frac{2i-1}{2n} \right)^2. \quad (4.20)$$

Dessa forma, sob \mathbf{H}_0 , a função de distribuição assintótica dessa estatística é (CSÖRGŐ; FARAWAY, 1996)

$$\lim_{n \rightarrow \infty} P(C_n \leq x) = 1 - \frac{2}{\pi} \sum_{i=1}^{\infty} (-1)^{i+1} \int_{(2i-1)\pi}^{2i\pi} \frac{e^{-t^2 \frac{x}{2}}}{\sqrt{-t \sin(t)}} dt.$$

4.5.4 Teste de Anderson-Darling

O teste Anderson-Darling (TAD) para uma amostra, foi proposto Anderson e Darling (1954) e também é utilizado para testar adequação de distribuições e as hipóteses do teste de são as mesmas descritas em (4.18). Sejam X_1, \dots, X_n uma amostra aleatória, considere ainda a amostra ordenada $X_{(1)}, \dots, X_{(n)}$, então a estatística do TAD é

$$A_n = -n - \frac{1}{n} \sum_{j=1}^n (2j-1) \{ \log[F(x_{(j)})] + \log[1 - F(x_{(n-j+1)})] \} \quad (4.21)$$

Dessa forma, sob \mathbf{H}_0 , a função de distribuição assintótica dessa estatística é (ANDERSON; DARLING, 1954)

$$\lim_{n \rightarrow \infty} P(A_n \leq x) = \frac{\sqrt{2\pi}}{x} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(j + \frac{1}{2})(4j + 1)}{j!} e^{-\frac{(4j+1)^2 \pi^2}{8x}} \int_0^{\infty} e^{\frac{x}{8(1+w^2)} - w^2(4j+1)^2 \pi^2 / 8x} dw.$$

Essa distribuição é bastante complicada, assim, Anderson e Darling (1954) fornecem os pontos de significância de 1%, 5% e 10% para grandes amostras.

4.5.5 Método *Bootstrap*

Na Estatística, o *bootstrap* é um método de reamostragem que foi proposto por Efron (1979). A ideia do método é gerar amostras artificiais que podem ser obtidas por amostragem a partir da amostra original como se estivéssemos amostrando da população (CORDEIRO; CRIBARI-NETO, 2014, p. 9). Esse mecanismo de amostragem pode ser feito de maneira paramétrica ou não paramétrica. O *bootstrap* permite obter medidas de interesse sem a necessidade de cálculos analíticos complicados. Além do mais, permite realizar correção de viés em estimadores e no melhoramento em testes de hipóteses. O interesse é utilizar o *bootstrap* para encontrar as distribuições empíricas dos testes: TRV, TKS, TCM e TAD, calcular o quantil correspondente ao tamanho α , ou ainda, o valor- p do teste.

Os passos para a utilização do *bootstrap* em teste de hipóteses estão descritos no Algoritmo 15.

Algoritmo 15: Teste de hipóteses por meio do método *bootstrap*.

Entrada: Amostra $\{x_i\}_{i=1}^n$ e o número B de interações.

Saída: Valor crítico do teste.

- 1:** Calcular o valor da estatística de teste denotada por $q(\mathbf{x})$ (utilize (4.17), (4.19), (4.20) ou (4.21));
 - 2:** Gere n pseudo-amostras aleatórias $\mathbf{x}_b = (x_1^b, \dots, x_n^b)$ sob a distribuição imposta em \mathbf{H}_0 ;
 - 3:** Calcule a estatística de teste com as amostras do passo 2, ou seja, $q^b = q^b(\mathbf{x}^b)$;
 - 4:** Repetir os passos 2 e 3 um número B grande de vezes;
 - 5:** Calcule o percentil estimado $\hat{q}_{(1-\alpha)}$ de interesse, utilizando a distribuição empírica das B realizações da estatística de teste $q^b(\mathbf{x}^b)$;
 - 6:** Faça o teste utilizando a estatística de teste da amostra original $q(\mathbf{x})$, calculada no passo 1, com o valor crítico *bootstrap* $\hat{q}_{(1-\alpha)}$, obtido no passo 5.
-

O teste *bootstrap* consiste em rejeitar \mathbf{H}_0 se $q(\mathbf{x}) > \widehat{q}_{(1-\alpha)}$. Ademais, pode-se calcular o valor- p *bootstrap* dado por

$$p^* = \frac{\sum_{b=1}^B I_{\{q^b > q(\mathbf{x})\}}}{B},$$

no qual

$$I_{\{q^b > q(\mathbf{x})\}} = \begin{cases} 1, & \text{se } q^b > q(\mathbf{x}), \\ 0, & \text{caso contrário.} \end{cases}$$

Portanto, \mathbf{H}_0 é rejeitada se p^* é menor que o valor α estabelecido. Portanto, nessa abordagem podem ser utilizadas as estatísticas de teste (4.16), (4.19), (4.20) e (4.21), não sendo necessário conhecer a distribuição de cada uma delas.

4.6 APROXIMAÇÃO DA DISTRIBUIÇÃO WATSON COMPLEXA PELA NORMAL MULTIVARIADA COMPLEXA SINGULAR

A distribuição Watson complexa foi definida na subseção 3.2.3 e algumas características foram estudadas. O objetivo é reescrever a densidade (3.5) de tal maneira, que se possa utilizar uma aproximação pela distribuição normal multivariada complexa singular, consequentemente, ocasionando uma aproximação em uma distribuição normal multivariada singular.

Inicialmente, considere a constante de normalização da distribuição Watson complexa, que é dada por

$$\begin{aligned} c(\kappa) &= 2\pi^{k-1} \kappa^{2-k} \left(e^\kappa - \sum_{j=0}^{k-3} \frac{\kappa^j}{j!} \right), \\ &= 2\pi^{k-1} \kappa^{2-k} e^\kappa \left(1 - \sum_{j=0}^{k-3} \frac{\kappa^j e^{-\kappa}}{j!} \right), \text{ se } \kappa \text{ é grande, então} \\ &\cong 2\pi^{k-1} \kappa^{2-k} e^\kappa. \end{aligned}$$

Assim,

$$f(\mathbf{z}) = \frac{e^{\kappa|\mathbf{z}^* \boldsymbol{\mu}|^2}}{2\pi^{k-1} \kappa^{2-k} e^\kappa} = \frac{e^{\kappa|\mathbf{z}^* \boldsymbol{\mu}|^2 - \kappa}}{2\pi^{k-1} \kappa^{2-k}} = \frac{e^{-\kappa(1-|\mathbf{z}^* \boldsymbol{\mu}|^2)}}{2\pi^{k-1} \kappa^{2-k}}.$$

Note que $|\mathbf{z}^* \boldsymbol{\mu}|^2 = \mathbf{z}^* \boldsymbol{\mu} \boldsymbol{\mu}^* \mathbf{z}$, então a densidade pode ser reescrita como

$$f(\mathbf{z}) = \frac{e^{-\kappa(1-\mathbf{z}^*\boldsymbol{\mu}\boldsymbol{\mu}^*\mathbf{z})}}{2\pi^{k-1}\kappa^{2-k}}.$$

Observe que,

$$(\boldsymbol{\mu} - \mathbf{z})^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)(\boldsymbol{\mu} - \mathbf{z}) = 1 - \mathbf{z}^*\boldsymbol{\mu}\boldsymbol{\mu}^*\mathbf{z}.$$

De fato, veja que $\mathbf{z}^*\mathbf{z} = \boldsymbol{\mu}^*\boldsymbol{\mu} = 1$, o que implica em

$$\begin{aligned} (\mathbf{z} - \boldsymbol{\mu})^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}) &= \boldsymbol{\mu}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\boldsymbol{\mu} + \mathbf{z}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\mathbf{z} - \mathbf{z}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\boldsymbol{\mu} - \boldsymbol{\mu}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\mathbf{z}, \\ &= \mathbf{z}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\mathbf{z}, \\ &= \mathbf{z}^*\mathbf{I}\mathbf{z} - \mathbf{z}^*\boldsymbol{\mu}\boldsymbol{\mu}^*\mathbf{z}, \\ &= 1 - \mathbf{z}^*\boldsymbol{\mu}\boldsymbol{\mu}^*\mathbf{z}. \end{aligned}$$

Definição 7. A inversa generalizada de uma matriz \mathbf{A} , diga-se \mathbf{A}^- , é uma matriz que satisfaz a condição de que $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

Ademais, $\mathbf{A}^- = \mathbf{A}$ em que $\mathbf{A} = (\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ e \mathbf{A}^- é a inversa generalizada de \mathbf{A} . Isso ocorre pelo fato de que, \mathbf{A} é uma matriz idempotente, ou seja, $\mathbf{A}^2 = \mathbf{A}$. De fato, veja que

$$\mathbf{A}^2 = (\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) = \mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^* - \boldsymbol{\mu}\boldsymbol{\mu}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) = \mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*.$$

Portanto, para κ grande, a densidade da distribuição Watson complexa pode ser aproximada da seguinte forma,

$$\begin{aligned} f(\mathbf{z}) &= \frac{e^{-\kappa(\mathbf{z}-\boldsymbol{\mu})^*(\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^*)^-(\mathbf{z}-\boldsymbol{\mu})}}{2\pi^{k-1}\kappa^{2-k}} \\ &= \frac{e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^*2\kappa(\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^*)^-(\mathbf{z}-\boldsymbol{\mu})}}{2\pi^{k-1}\kappa^{2-k}} \\ &= \frac{e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^*2\kappa(\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^*)^-(\mathbf{z}-\boldsymbol{\mu})}}{2^{1+2+k-k-2}\pi^{k-1}\kappa^{2-k}} \\ &= \frac{e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^*2\kappa(\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^*)^-(\mathbf{z}-\boldsymbol{\mu})}}{(2\pi)^{k-1}(2\kappa)^{2-k}}. \end{aligned}$$

A matriz $(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ é uma matriz $(k-1) \times (k-1)$ hermitiana positiva semi-definida e idempotente. Assim, pode-se verificar que o posto de $(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ é igual a $k-2$. De fato, seja \mathbf{U} uma matriz $(k-1) \times (k-1)$ ortogonal, tal que

$$|\mathbf{U}| = 1.$$

Considere também que $\boldsymbol{\mu}_u = \mathbf{U}\boldsymbol{\mu} = (1, 0, \dots, 0)^\top$ com $k-2$ zeros. Assim, encontremos os autovalores de $(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ resolvendo o sistema

$$\left| \frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) - \lambda\mathbf{I} \right| = 0,$$

ou, equivalentemente

$$\left| \mathbf{U} \left[\frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) - \lambda\mathbf{I} \right] \mathbf{U}^* \right| = 0.$$

Observe que

$$\begin{aligned} \mathbf{U} \left[\frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) - \lambda\mathbf{I} \right] \mathbf{U}^* &= \left[\frac{1}{2\kappa}(\mathbf{I} - \mathbf{U}\boldsymbol{\mu}\boldsymbol{\mu}^*\mathbf{U}^*) - \lambda\mathbf{I} \right] \\ &= \left[\frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}_u\boldsymbol{\mu}_u^*) - \lambda\mathbf{I} \right]. \end{aligned}$$

Além disso,

$$\boldsymbol{\mu}_u\boldsymbol{\mu}_u^* = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Portanto,

$$\begin{aligned} \mathbf{U} \left[\frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) - \lambda\mathbf{I} \right] \mathbf{U}^* &= \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & (2\kappa)^{-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (2\kappa)^{-1} \end{pmatrix} - \lambda\mathbf{I} \\ &= \begin{pmatrix} -\lambda & 0 & 0 & \cdots & 0 \\ 0 & (2\kappa)^{-1} - \lambda & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (2\kappa)^{-1} - \lambda \end{pmatrix}. \end{aligned}$$

Assim, temos uma matriz diagonal e o seu determinante é igual ao produto dos elementos da diagonal principal. Assim sendo,

$$\left| \frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*) - \lambda\mathbf{I} \right| = -\lambda[(2\kappa)^{-1} - \lambda]^{k-2}.$$

Desse modo, temos o seguinte sistema para resolver

$$-\lambda[(2\kappa)^{-1} - \lambda]^{k-2} = 0.$$

É fácil ver que as soluções desse sistema são $\lambda_i = (2\kappa)^{-1}$ para $i = 1, 2, \dots, k-2$ e $\lambda_{k-1} = 0$. Portanto, como $\lambda_i > 0$ ($i = 1, 2, \dots, k-2$) temos que o posto de $(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ é igual a $k-2$ e

$$\prod_{i=1}^{k-2} \lambda_i = (2\kappa)^{-(k-2)} = (2\kappa)^{2-k}.$$

Consequentemente, pode-se ver que $\mathbf{Z} \sim \mathbb{CSN}_{k-1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com $\boldsymbol{\Sigma} = \frac{1}{2\kappa}(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$. Utilizando o resultado (4.15), tem-se que

$$2\kappa(\mathbf{Z} - \boldsymbol{\mu})^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)^{-1}(\mathbf{Z} - \boldsymbol{\mu}) = 2\kappa\mathbf{Z}^*(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)^{-1}\mathbf{Z} \approx \chi_{2(k-2)}^2. \quad (4.22)$$

Esse resultado foi proposto por Mardia e Dryden (1999) e é utilizado para realizar testes de hipóteses. Entretanto, mostramos passos que complementam essa aproximação que não foram demonstrados em Mardia e Dryden (1999), tais como a aproximação da Watson complexa para a normal multivariada singular complexa quando κ é grande e o resultado da equação (4.15).

Perceba que esse resultado é bastante interessante, pois independe do valor de κ e do vetor de médias $\boldsymbol{\mu}$. Ou seja, a distribuição resultante não depende dos parâmetros da distribuição Watson complexa, depende apenas da dimensão de $\boldsymbol{\mu}$. Ademais, nenhuma análise foi feita para verificar a qualidade dessa aproximação. A seguir, foi feito um estudo de simulação considerando várias configurações para avaliar esse resultado com o intuito de identificar o valor mínimo de κ para que a aproximação seja válida.

4.6.1 Avaliação da Aproximação para a Distribuição Qui-quadrado - Watson Complexa

Foram geradas $n = 200$ amostras da distribuição $\mathbb{CW}_{6-2}(\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top, \kappa = 100)$, e utilizando o resultado (4.22), obtém-se uma distribuição χ_8^2 . Na Figura 25, estão duas análises gráficas para o resultado e, pode-se ver em ambos os gráficos, que o ajuste da distribuição χ_8^2 é excelente.

Para ampliar esse resultado, serão utilizadas 5000 réplicas de Monte Carlo para averiguar o tamanho do TRV, TKS, TCM e TAD para diferentes configurações dos parâmetros e número de amostras. Os valores utilizados foram:

- A forma média foi fixada em $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$;
- Para $\kappa = 1, 3, 5, 8, 10, 20, 30, 40, 50, 100, 1000, 5000$;
- Com $n = 20, 30, 40, 50, 100, 200$.

Na Tabela 9, estão os resultados da simulação utilizando um tamanho de $\alpha = 5\%$. Nota-se que para os valores de $\kappa \geq 10$, as taxas de rejeição se aproximam do nível de 5%, para todos os valores de n . Portanto, conclui-se que os testes funcionaram muito bem. No Apêndice A, nas Tabelas 21 e 22 estão as taxas de rejeição para a simulação utilizando $\alpha = 1\%$ e 10%, respectivamente. A mesma situação ocorre, indicando que os testes também funcionaram para $\kappa \geq 10$. Best e Fisher (1986) concluíram as mesmas coisas em relação aos valores de κ , porém para a distribuição Watson na esfera real.

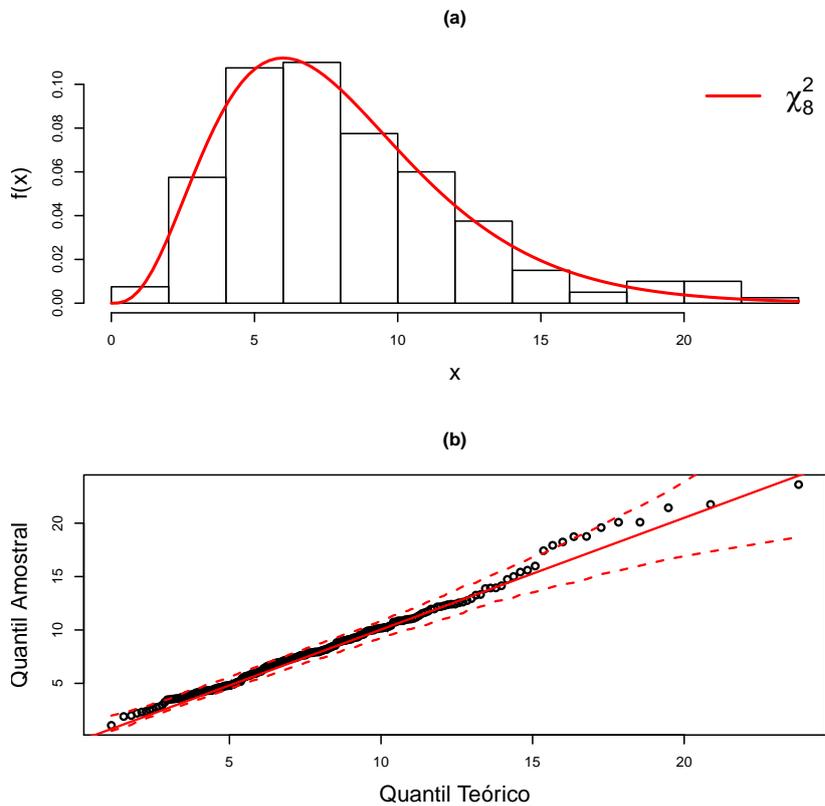


Figura 25 – Análise gráfica utilizando a aproximação (4.22), utilizando dados da distribuição Watson complexa com $\kappa = 100$ e $n = 200$. Gráfico (a): Histograma; (b): Quantil-Quantil.

Tabela 9 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\mu = (0, 0, 0, 0, 1)^T$, diferentes tamanhos amostrais e utilizando o nível de significância de 5%.

Teste	n	κ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	1,0000	0,4024	0,0478	0,0432	0,0492	0,0522	0,0450	0,0468	0,0494	0,0510	0,0476
	30	1,0000	1,0000	0,6182	0,0618	0,0454	0,0552	0,0500	0,0514	0,0566	0,0528	0,0422	0,0478
	40	1,0000	1,0000	0,7806	0,0642	0,0486	0,0474	0,0510	0,0484	0,0502	0,0528	0,0508	0,0446
	50	1,0000	1,0000	0,8892	0,0698	0,0522	0,0486	0,0532	0,0500	0,0430	0,0434	0,0472	0,0540
	100	1,0000	1,0000	0,9990	0,0986	0,0428	0,0526	0,0458	0,0474	0,0526	0,0494	0,0448	0,0460
	200	1,0000	1,0000	1,0000	0,1844	0,0588	0,0430	0,0544	0,0484	0,0538	0,0496	0,0474	0,0434
TKS	20	1,0000	1,0000	0,6396	0,0582	0,0456	0,0466	0,0518	0,0458	0,0498	0,0504	0,0518	0,0478
	30	1,0000	1,0000	1,0000	0,0606	0,0482	0,0494	0,0500	0,0522	0,0502	0,0524	0,0444	0,0508
	40	1,0000	1,0000	1,0000	0,0618	0,0490	0,0440	0,0488	0,0486	0,0522	0,0442	0,0450	0,0484
	50	1,0000	1,0000	1,0000	0,0580	0,0494	0,0502	0,0524	0,0516	0,0472	0,0468	0,0466	0,0538
	100	1,0000	1,0000	1,0000	0,0692	0,0424	0,0476	0,0400	0,0458	0,0450	0,0446	0,0424	0,0426
	200	1,0000	1,0000	1,0000	0,1210	0,0496	0,0452	0,0478	0,0506	0,0506	0,0438	0,0488	0,0432
TCM	20	1,0000	1,0000	0,6182	0,0544	0,0486	0,0460	0,0526	0,0488	0,0522	0,0512	0,0484	0,0462
	30	1,0000	1,0000	0,9002	0,0624	0,0468	0,0490	0,0488	0,0518	0,0542	0,0512	0,0436	0,0448
	40	1,0000	1,0000	0,9908	0,0672	0,0478	0,0472	0,0480	0,0478	0,0498	0,0498	0,0502	0,0454
	50	1,0000	1,0000	1,0000	0,0676	0,0516	0,0500	0,0516	0,0514	0,0484	0,0464	0,0470	0,0514
	100	1,0000	1,0000	1,0000	0,0890	0,0440	0,0508	0,0452	0,0484	0,0526	0,0476	0,0452	0,0436
	200	1,0000	1,0000	1,0000	0,1564	0,0550	0,0454	0,0524	0,0530	0,0548	0,0448	0,0478	0,0448
TAD	20	1,0000	1,0000	0,6042	0,0494	0,0454	0,0488	0,0524	0,0470	0,0488	0,0506	0,0500	0,0458
	30	1,0000	1,0000	0,9230	0,0620	0,0460	0,0518	0,0494	0,0528	0,0548	0,0518	0,0442	0,0478
	40	1,0000	1,0000	0,9988	0,0664	0,0512	0,0468	0,0508	0,0482	0,0506	0,0484	0,0500	0,0430
	50	1,0000	1,0000	1,0000	0,0712	0,0510	0,0482	0,0484	0,0532	0,0452	0,0466	0,0470	0,0530
	100	1,0000	1,0000	1,0000	0,1022	0,0432	0,0522	0,0448	0,0474	0,0564	0,0476	0,0468	0,0452
	200	1,0000	1,0000	1,0000	0,2272	0,0560	0,0452	0,0520	0,0516	0,0526	0,0464	0,0478	0,0474

Na prática, os parâmetros da distribuição Watson complexa não são conhecidos e precisam ser estimados (ver a estimação na subseção 4.2). Assim, a ideia é utilizar o resultado (4.22) com as estimativas dos parâmetros via MV, como feito em Fisher e Best (1984) para a distribuição de Fisher e em Best e Fisher (1986) para a distribuição Watson, ambas na esfera real. Nos dois casos, foram utilizados o TKS e o teste de Kuiper com correções nas estatísticas de teste, fornecidas por Stephens (1974). Entretanto, essas correções foram propostas apenas quando os dados possuem distribuição normal ou exponencial. Além disso, Johnson e Wichern (2007, p. 184-185) utilizam os estimadores de MV para os parâmetros da distribuição normal multivariada, no resultado da equação (4.10), e utilizam um gráfico, chamado de gráfico qui-quadrado ou gráfico gama, para averiguar a normalidade multivariada em um conjunto de dados.

Os testes foram utilizados com os parâmetros estimados, porém os tamanhos dos testes ficaram muito abaixo do valor verdadeiro. O método *bootstrap* foi utilizado em todos os testes, mesmo assim, também não produziram melhoras. Dessa forma, foi proposta uma transformação de variável no resultado da equação (4.22), que recai em uma variável aleatória beta, em que não será necessário estimar o parâmetro κ e, conseqüentemente, percebeu-se uma melhora no desempenho dos testes.

4.6.2 Distribuição de Dirichlet

A distribuição de Dirichlet é uma família de distribuições de probabilidade multivariadas contínuas parametrizadas por um vetor de parâmetros positivos. Essa família pode ser vista como uma generalização multivariada da distribuição Beta. A distribuição de Dirichlet é muito utilizada em inferência Bayesiana, devido ao fato de que a distribuição *a posteriori* também segue uma distribuição de Dirichlet. Seja \mathbf{X} um vetor aleatório com distribuição Dirichlet, diga-se $\mathbf{X} \sim Dr(\mathbf{a})$, em que $\mathbf{a} = (a_1, \dots, a_p)^\top$. Então, sua densidade é dada por

$$f(\mathbf{x}) = \frac{\Gamma(a_1 + \dots + a_p)}{\Gamma(a_1) \dots \Gamma(a_p)} \prod_{i=1}^p x_i^{a_i-1}, \quad (4.23)$$

no qual \mathbf{x} assume valores no *simplex* $\mathbb{S} = \{(x_1, \dots, x_p) : x_p > 0, \sum_{i=1}^p x_i = 1\}$, no qual cada $a_i > 0 (i = 1, \dots, p)$. Perceba que, se $p = 2$, então \mathbf{X} tem distribuição *Beta*(a_1, a_2), que possui a densidade

$$f(\mathbf{x}) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} x_1^{a_1-1} (1-x_2)^{a_2-1}, \quad x_i \in (0, 1).$$

Muitas vezes a densidade (4.23) é reescrita para

$$f(\mathbf{x}) = \frac{\Gamma(a_1 + \dots + a_p)}{\Gamma(a_1) \dots \Gamma(a_p)} \left(\prod_{i=1}^{p-1} x_i^{a_i-1} \right) \left(1 - \sum_{i=1}^{p-1} x_i \right)^{a_p-1},$$

assumindo valores no *simplex* $\mathbb{S} = \{(x_1, \dots, x_{p-1}) : x_i > 0, \sum_{i=1}^{p-1} x_i < 1\}$, utilizando o fato de que $x_p = 1 - \sum_{i=1}^{p-1} x_i$. Além disso, essa distribuição possui as seguintes propriedades (FERGUSON, 1973)

$$\begin{aligned} E(X_i) &= \frac{a_i}{\sum_{i=1}^p a_i} \\ \text{Var}(X_i) &= \frac{a_i(\sum_{i=1}^p a_i - a_i)}{(\sum_{i=1}^p a_i)^2(\sum_{i=1}^p a_i + 1)} \\ \text{Cov}(X_i, X_j) &= \frac{-a_i a_j}{(\sum_{i=1}^p a_i)^2(\sum_{i=1}^p a_i + 1)} \quad \text{para } i \neq j. \end{aligned}$$

Além do mais, a distribuição marginal para cada X_i é $Beta(a_i, \sum_{j=1}^n a_j - a_i)$.

Existe uma relação interessante com a distribuição Gama, como será visto a seguir. Seja X uma variável aleatória com distribuição $Gama(a, b)$, então a densidade de X é dada por

$$f(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}}, \quad 0 < x < \infty,$$

para $a > 0$ e $b > 0$. Agora, considere uma amostra aleatória X_1, \dots, X_n independentes, no qual $X_i \sim Gama(a_i, b)$, para $i = 1, \dots, n$. A distribuição conjunta é dada por,

$$\begin{aligned} f(\mathbf{x}) &= \prod_{i=1}^n \frac{1}{\Gamma(a_i)b^{a_i}} x_i^{a_i-1} e^{-\frac{x_i}{b}} \\ &= \left(\prod_{i=1}^n \frac{1}{\Gamma(a_i)b^{a_i}} \right) \left(\prod_{i=1}^n x_i^{a_i-1} \right) e^{-\frac{\sum_{i=1}^n x_i}{b}}. \end{aligned}$$

Teorema 5. Considere uma amostra aleatória X_1, \dots, X_n de $X \sim Gama(a_i, b)$. Seja $Y_j = X_j / \sum_{i=1}^n X_i$, para $j = 1, \dots, n-1$. Então, $\mathbf{Y} = (Y_1, \dots, Y_{n-1})^\top \sim Dr(a_1, \dots, a_n)$.

Para verificar a prova do Teorema 5, ver Devroye (1986, p. 594). Portanto, pode-se utilizar o resultado do Teorema 5 no resultado da equação (4.22), pois existe a seguinte relação: se $X \sim \chi_v^2$, então $X \sim Gama(\frac{v}{2}, \frac{1}{2})$. Considere que

$$V_i = 2\kappa \mathbf{Z}_i^* (\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}^*)^{-1} \mathbf{Z}_i, \text{ para } i = 1, \dots, n.$$

Então, $V_i \approx \chi_{2k-4}^2$, equivalentemente, $V_i \approx Gama(k-2, \frac{1}{2})$. À vista disso, fazendo

$$\begin{aligned} Y_j &= \frac{V_j}{\sum_{i=1}^n V_i} = \frac{2\kappa \mathbf{Z}_j^* (\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}^*)^{-1} \mathbf{Z}_j}{\sum_{i=1}^n 2\kappa \mathbf{Z}_i^* (\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}^*)^{-1} \mathbf{Z}_i} \\ &= \frac{\mathbf{Z}_j^* (\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}^*)^{-1} \mathbf{Z}_j}{\sum_{i=1}^n \mathbf{Z}_i^* (\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}^*)^{-1} \mathbf{Z}_i}. \end{aligned} \quad (4.24)$$

Desse modo, Y_1, \dots, Y_{n-1} possui distribuição $Dr(a_1, \dots, a_n)$, em que $a_i = k - 2 (i = 1, \dots, n)$. Note que, utilizando essa abordagem não há necessidade de se conhecer o valor de κ . Quando os parâmetros da distribuição de Dirichlet são todos iguais, essa distribuição recebe o nome de distribuição simétrica de Dirichlet.

A densidade conjunta de (4.24) é uma distribuição simétrica de Dirichlet, e possui a densidade

$$f(y_1, \dots, y_{n-1}) = \frac{\Gamma(na)}{\Gamma(a)^n} \left(\prod_{i=1}^{n-1} y_i^{a-1} \right) \left(1 - \sum_{i=1}^{n-1} y_i \right)^{a-1},$$

para $(y_1, \dots, y_{n-1}) \in \mathbb{S}$ e $a = k - 2$. A distribuição marginal para cada Y_i é beta com densidade dada por

$$f(y_i) = \frac{\Gamma(na)}{\Gamma(a)\Gamma((n-1)a)} y_i^{a-1} (1 - y_i)^{(n-1)a-1}, \quad y_i \in (0, 1).$$

Contudo, aparece um problema em utilizar essa abordagem, pois com essa transformação, as variáveis não são independentes. Dessa forma, os testes TKS, TCM e TAD não poderão ser utilizados, pois não existem versões desses testes para variáveis correlacionadas. Dessa forma, será utilizada a transformação (4.24) com os parâmetros estimados e o método *bootstrap* para estimar as distribuições das quatro estatísticas de teste empiricamente.

4.7 AVALIAÇÃO NUMÉRICA

4.7.1 Tamanho do Teste

Na Tabelas 10 estão os resultados das simulações utilizando a transformação da equação (4.24) com a forma média estimada, considerando 5000 réplicas de Monte Carlo, considerando a forma média fixada em $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, para $\kappa = 10, 50, 5000$, utilizando $n = 20, 30, 40, 50, 100, 200$ e considerando os tamanhos $\alpha = 1\%, 5\%, 10\%$ para os testes TRV, TKS, TCM e TAD. Pode-se ver que para o valor de $\kappa = 10$, os tamanhos dos quatro testes estão muito próximos do valor teórico, indicando que os testes de adequação produziram um bom desempenho. Além disso, percebe-se que o desempenho dos testes não muda quando o número de amostra cresce. Outro ponto interessante, é que para valores pequenos de n e altos valores de κ as taxas de rejeição também se aproxima do valor nominal verdadeiro. Para verificar outras configurações, ver as Tabelas 23, 24 e 25 no Apêndice A.

Tabela 10 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, considerando a forma média $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$, para valores de $\kappa = 10, 50, 5000$, tamanhos amostrais $n = 20, 30, 40, 50, 100, 200$ e utilizando $\alpha = 1\%, 5\%, 10\%$.

Teste	κ α	10			50			5000			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
TRV	n	20	0,0088	0,0478	0,0942	0,0106	0,0536	0,1024	0,0124	0,0506	0,1010
		30	0,0096	0,0458	0,0892	0,0126	0,0514	0,1012	0,0108	0,0484	0,0986
		40	0,0066	0,0564	0,0874	0,0114	0,0504	0,0948	0,0090	0,0566	0,1062
		50	0,0096	0,0504	0,0938	0,0096	0,0494	0,0992	0,0108	0,0518	0,1018
		100	0,0070	0,0520	0,1054	0,0084	0,0448	0,0924	0,0104	0,0496	0,1082
		200	0,0160	0,0458	0,1268	0,0100	0,0496	0,0958	0,0068	0,0442	0,0956
TKS	n	20	0,0116	0,0518	0,0902	0,0106	0,0488	0,0998	0,0110	0,0484	0,1022
		30	0,0072	0,0504	0,0864	0,0122	0,0528	0,1060	0,0100	0,0470	0,0962
		40	0,0082	0,0522	0,0924	0,0118	0,0518	0,1002	0,0114	0,0540	0,1040
		50	0,0078	0,0476	0,0934	0,0112	0,0484	0,0988	0,0124	0,0538	0,1040
		100	0,0082	0,0498	0,0918	0,0106	0,0540	0,1038	0,0106	0,0540	0,1034
		200	0,0134	0,0462	0,1156	0,0086	0,0442	0,0986	0,0090	0,0516	0,1048
TCM	n	20	0,0110	0,0550	0,0934	0,0094	0,0534	0,1028	0,0098	0,0524	0,1022
		30	0,0092	0,0504	0,0884	0,0128	0,0566	0,1024	0,0088	0,0504	0,1032
		40	0,0084	0,0534	0,0908	0,0114	0,0538	0,1022	0,0100	0,0530	0,1084
		50	0,0096	0,0484	0,0916	0,0094	0,0506	0,0942	0,0112	0,0528	0,1056
		100	0,0066	0,0504	0,0910	0,0110	0,0586	0,1072	0,0104	0,0500	0,1086
		200	0,0126	0,0478	0,1150	0,0094	0,0476	0,1010	0,0076	0,0492	0,0990
TAD	n	20	0,0120	0,0588	0,1000	0,0108	0,0506	0,0990	0,0116	0,0518	0,1044
		30	0,0100	0,0522	0,0880	0,0114	0,0544	0,1048	0,0090	0,0484	0,1004
		40	0,0072	0,0562	0,0878	0,0126	0,0510	0,1036	0,0100	0,0534	0,1098
		50	0,0104	0,0568	0,0948	0,0106	0,0478	0,0948	0,0120	0,0544	0,1082
		100	0,0054	0,0494	0,0892	0,0102	0,0564	0,1074	0,0104	0,0516	0,1034
		200	0,0118	0,0488	0,1110	0,0092	0,0498	0,1010	0,0068	0,0478	0,0988

4.7.2 Poder do Teste

Para a análise do poder foi utilizado um mecanismo de contaminação na amostra, como feito em Fisher e Best (1984) para a distribuição de Fisher na esfera real. Visto que existem muitos parâmetros para serem avaliados da distribuição Watson. O mecanismo se baseia na contaminação em 10% do tamanho amostral, em que essas amostras contaminadas serão geradas de outra variável com distribuição Watson com uma forma média diferente. O esquema de avaliação é descrito a seguir:

- A forma média encontrar-se-á fixada em $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$;
- As formas médias contaminadas são geradas somando um valor real na forma média original. Serão utilizados sete valores de contaminação e eles são descritos no vetor $\boldsymbol{\varepsilon}$ dado por

$$\boldsymbol{\varepsilon} = \begin{pmatrix} 0,0357 \\ 0,0714 \\ 0,1071 \\ 0,1429 \\ 0,1786 \\ 0,2143 \\ 0,2500 \end{pmatrix}. \quad (4.25)$$

Assim, gerando as formas médias contaminadas da seguinte forma

$$\boldsymbol{\mu}_i = \frac{\boldsymbol{\mu} + \boldsymbol{\varepsilon}_i}{\|\boldsymbol{\mu} + \boldsymbol{\varepsilon}_i\|}, \quad i = 1, 2, \dots, 7, \quad (4.26)$$

no qual $\boldsymbol{\varepsilon}_i$ é o i -ésimo elemento de $\boldsymbol{\varepsilon}$. Na Tabela 11 estão os valores das formas médias geradas com os respectivos valores de contaminação. Pode-se calcular a distância geodésica entre a forma média original e cada forma média resultante com a contaminação, resultando no vetor

$$\boldsymbol{\theta} = \begin{pmatrix} 0,07 \\ 0,13 \\ 0,19 \\ 0,24 \\ 0,29 \\ 0,34 \\ 0,38 \end{pmatrix},$$

Tabela 11 – Resumo das formas médias utilizadas na análise de poder com o respectivo valor de contaminação.

Forma média	ϵ	Coordenadas				
μ_1	0,0357	0,0344	0,0344	0,0344	0,0344	0,9976
μ_2	0,0714	0,0661	0,0661	0,0661	0,0661	0,9912
μ_3	0,1071	0,0950	0,0950	0,0950	0,0950	0,9818
μ_4	0,1429	0,1213	0,1213	0,1213	0,1213	0,9701
μ_5	0,1786	0,1450	0,1450	0,1450	0,1450	0,9570
μ_6	0,2143	0,1664	0,1664	0,1664	0,1664	0,9430
μ_7	0,2500	0,1857	0,1857	0,1857	0,1857	0,9285

no qual θ_i corresponde ao i -ésimo elemento de θ e é a distância geodésica entre μ e μ_i .

- Com os valores de $\kappa = 10, 20, 30, 40, 50, 100, 1000, 5000$;
- Com o número de amostras $n = 20, 30, 40, 50, 100, 200$ e, conseqüentemente, o número de amostras contaminadas são 2, 3, 4, 5, 10, 20, respectivamente.

Primeiramente, considerando $\alpha = 5\%$, concluiu-se que para $\kappa \geq 30$ o poder começa a aumentar à medida que o valor da contaminação aumenta para todos os tamanhos amostrais. Entretanto, só quando $\kappa = 50$ com $n = 200$, que o poder chega à mais de 50% em todos os testes, porém quando $\kappa > 50$ o poder do teste aumenta consideravelmente para todos os números de amostras. Considerando $\kappa = 100$, o poder se torna atrativo, chegando em altos valores a partir de ϵ_6 como mostrado na Tabela 12. Ademais, para um dado tamanho de amostra n fixa, à medida que o valor da contaminação aumenta, o poder também aumenta.

Com o intuito de melhorar a visualização dos resultados da Tabela 12, observe na Figura 26 as curvas de poder para os testes de bondade de ajuste, considerando $n = 20, 40, 200$. Percebe-se que quanto maior o tamanho da amostra, maior o poder.

Na Figura 27 pode-se analisar o desempenho entre os testes em relação ao poder. Veja que, considerando $n = 20, 40$ o TAD obteve um desempenho maior do que os demais testes. Além do mais, quando o tamanho da amostra é muito grande a diferença entre os testes é pequena. Considerando uma concentração um pouco maior com $\kappa = 1000$, veja na Figura 28, e perceba que os desempenhos dos testes são praticamente os mesmos. As análises do poder também foram realizadas considerando $\alpha = 1\%$ e 10%, produzindo resultados similares (ver Apêndice A).

Tabela 12 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível de significância de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0524	0,0526	0,0550	0,0918	0,1930	0,3422	0,5156
	30	0,0504	0,0504	0,0688	0,1074	0,2578	0,4430	0,6768
	40	0,0506	0,0514	0,0620	0,1392	0,3114	0,5688	0,7898
	50	0,0492	0,0470	0,0694	0,1744	0,3778	0,6442	0,8662
	100	0,0532	0,0568	0,0932	0,2812	0,6182	0,9054	0,9878
	200	0,0484	0,0612	0,1390	0,4736	0,8924	0,9968	1,0000
TKS	20	0,0500	0,0566	0,0708	0,1062	0,2250	0,3678	0,5312
	30	0,0486	0,0482	0,0750	0,1298	0,2648	0,4594	0,6740
	40	0,0514	0,0512	0,0784	0,1414	0,3038	0,5488	0,7726
	50	0,0528	0,0498	0,0722	0,1632	0,3598	0,6250	0,8322
	100	0,0522	0,0552	0,0914	0,2482	0,5610	0,8708	0,9810
	200	0,0492	0,0562	0,1232	0,4140	0,8446	0,9920	1,0000
TCM	20	0,0508	0,0552	0,0696	0,1164	0,2500	0,4192	0,6094
	30	0,0506	0,0500	0,0750	0,1446	0,3070	0,5278	0,7674
	40	0,0500	0,0504	0,0834	0,1554	0,3532	0,6372	0,8548
	50	0,0544	0,0522	0,0760	0,1940	0,4232	0,7114	0,9078
	100	0,0542	0,0608	0,0980	0,2910	0,6678	0,9320	0,9962
	200	0,0494	0,0612	0,1344	0,4878	0,9180	0,9984	1,0000
TAD	20	0,0562	0,0628	0,0786	0,1454	0,3044	0,4868	0,6880
	30	0,0504	0,0558	0,0918	0,1714	0,3624	0,5912	0,8182
	40	0,0528	0,0570	0,0912	0,1898	0,4100	0,6992	0,8920
	50	0,0528	0,0548	0,0832	0,2316	0,4752	0,7692	0,9346
	100	0,0586	0,0632	0,1120	0,3346	0,7148	0,9542	0,9968
	200	0,0484	0,0674	0,1464	0,5338	0,9392	0,9996	1,0000

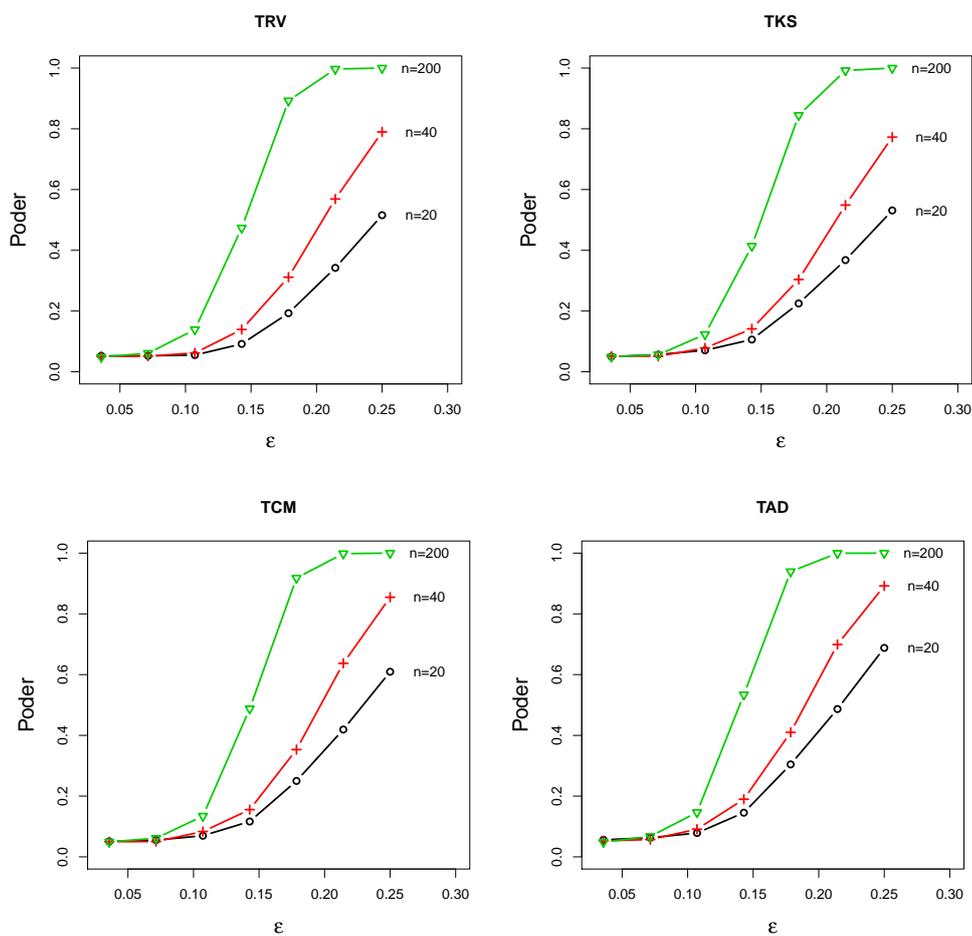


Figura 26 – Curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ε com 7 valores, $\kappa = 100$ e $n = 20, 40, 200$.

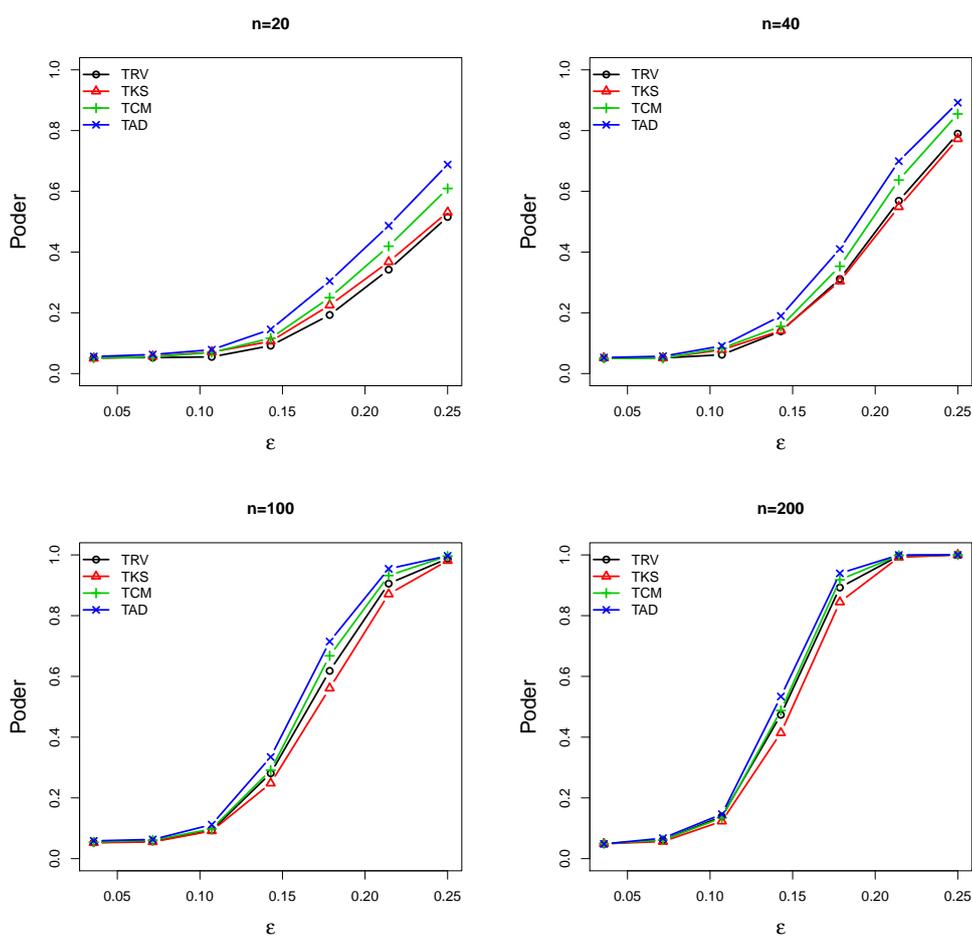


Figura 27 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ϵ com 7 valores, $\kappa = 100$ e $n = 20, 40, 100, 200$.

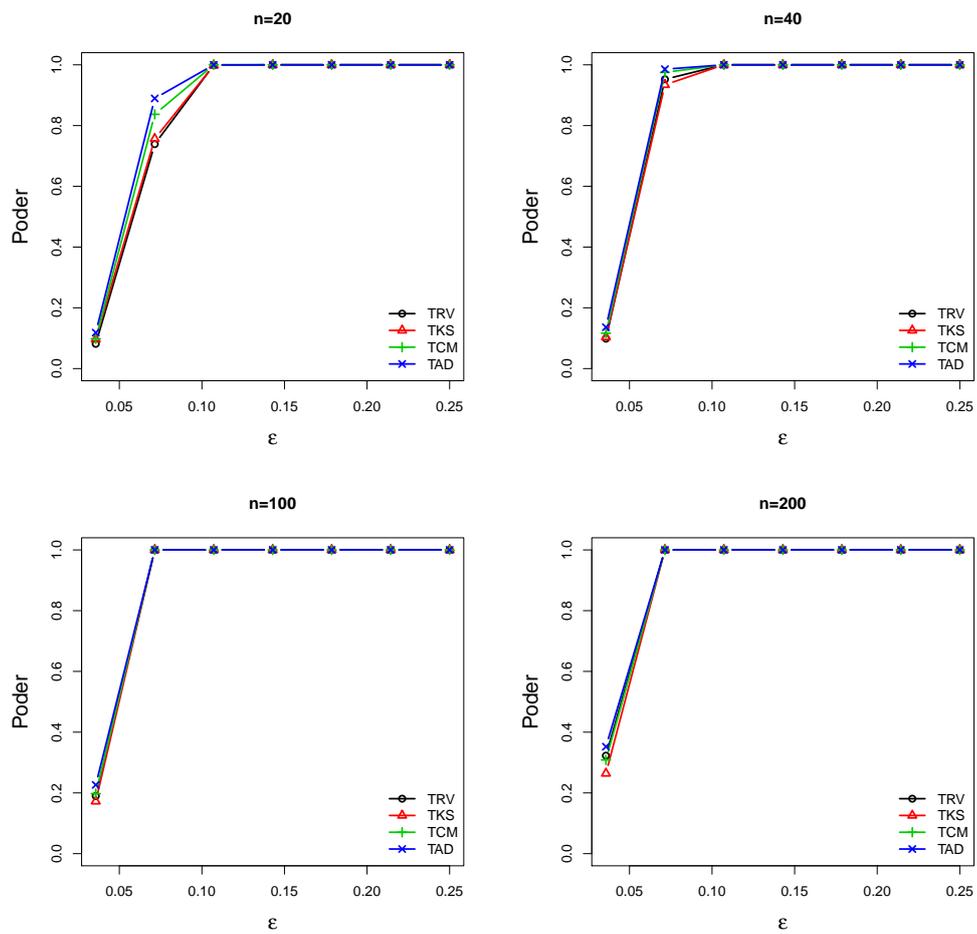


Figura 28 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ϵ com 7 valores, $\kappa = 1000$ e $n = 20, 40, 100, 200$.

4.8 APLICAÇÃO EM DADOS REAIS

Os dados utilizados são os mesmos discutidos na subseção 3.11 com a adição de novos dados que também podem ser encontrados no pacote `shapes` do R. Os novos bancos de dados são:

1. Crânios de Primatas: Chimpanzé

Como o objetivo de avaliar diferenças cranianas entre os sexos de primatas, foram coletadas amostras de crânios de chimpanzés adultos, que consistem em 26 fêmeas e 28 machos com 8 marcos. Na Figura 29 estão os macos nas coordenadas procrustes para ambos os sexos.

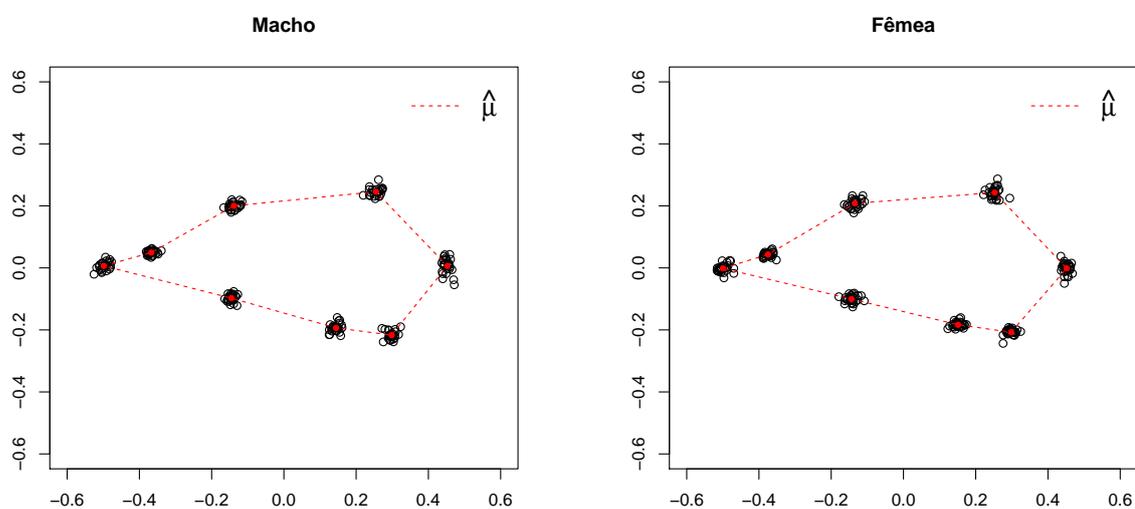


Figura 29 – Marcos nas coordenadas procrustes dos crânios de chimpanzés.

2. Conchas

Esse banco de dados possui 3 marcos tomados de 21 conchas. Esses dados foram publicados por Lohmann (1983) e foram baseados em amostras aleatórias de organismos tomados em diferentes latitudes no Oceano Índico Sul. Na Figura 30 estão os dados utilizando as coordenadas procrustes.

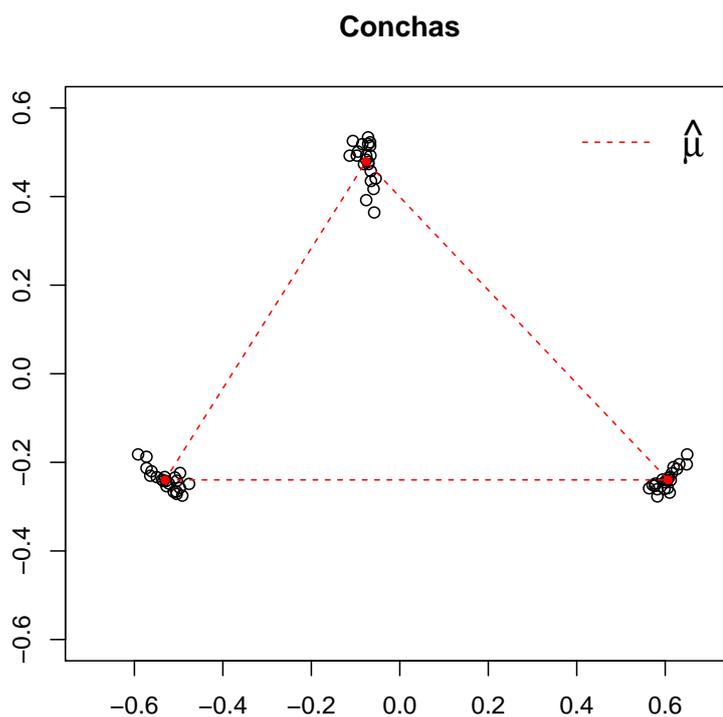


Figura 30 – Marcos nas coordenadas procrustes de microfósseis de conchas.

3. Dígito 3

Uma amostra aleatória de códigos postais feitos à mão de britânicos foi coletada e digitalizada por Anderson (1997). O interesse é examinar a forma e a variabilidade média, que podem ser usadas como um modelo anterior para reconhecimento de dígitos a partir de imagens de códigos postais manuscritos. Um dos dígitos de exemplo é o dígito 3. A amostra do dígito corresponde a 30 com 13 marcos e na Figura 31 pode-se observar esses dados nas coordenadas procrustes.

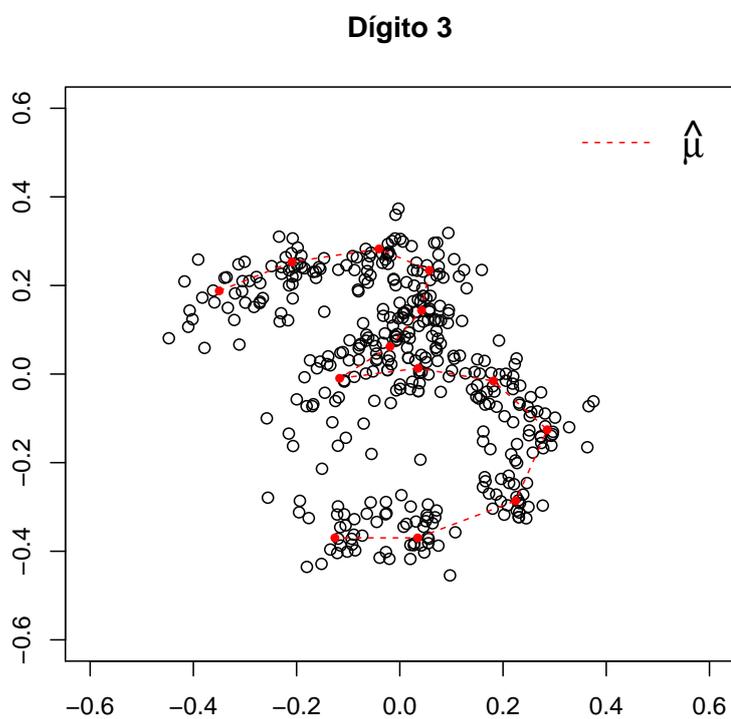


Figura 31 – Marcos nas coordenadas procrustes dos dígitos manuscritos.

4. Crânios de Ratos

Esse conjunto de dados foi extraído através de raio-X de crânios de 18 ratos. Os ratos foram cuidadosamente radiografados aos 7, 14, 21, 30, 40, 60, 90 e 150 dias, com 8 marcros em duas dimensões. Os dados estão descritos em Bookstein (1997, p. 67-69) e na Figura 32 estão presentes os dados referentes ao sétimo dia nas coordenadas procrustes.

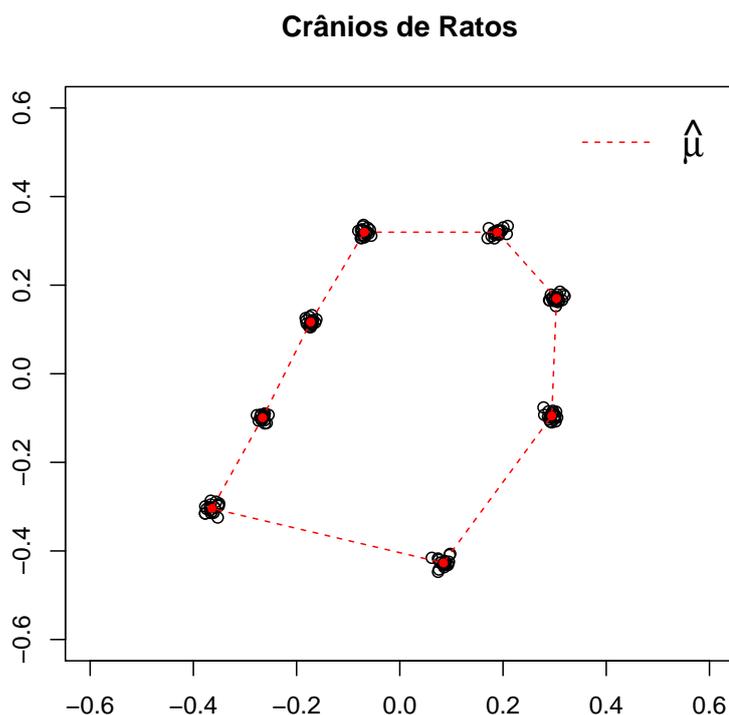


Figura 32 – Marcros nas coordenadas procrustes de crânios de ratos no sétimo dia.

Na Tabela 13 tem-se um resumo de alguns valores para os testes de adequação, como: o tamanho da amostra n , número de marcros k , estimativa de κ e os valores de significância para cada teste para os dados dos gorilas, orangotangos, vértebras, dígitos e conchas. Inicialmente, note que todos os valores estimados de κ são altos e o menor valor é $\hat{\kappa} = 146$ para os dados do dígito manuscrito. Indicando que se os dados possuírem distribuição Watson complexa, a aproximação para a distribuição normal multivariada complexa singular será boa e, conseqüentemente, os testes funcionarão.

Para tomar a decisão de rejeitar ou não se os dados provêm da distribuição Watson complexa, iremos adotar a decisão da maioria dos testes. Pela Tabela 13, pode-se ver que ao nível de significância de 5%, considerando a decisão dada pela maioria dos testes, que os dados referentes ao dígito 3, vértebras pequenas, gorilas e chipanzés para ambos os sexos, não possuem

distribuição Watson complexa. Já para o caso dos orangotangos fêmea, pode-se dizer que esses dados possuem distribuição Watson complexa. Em relação aos dados das conchas e para as vértebras grandes, houve uma divergência entre os testes ao nível de 5%, porém caso fosse considerado um nível de 1%, poderia-se dizer que esses dados também possuem distribuição Watson complexa. Outro ponto importante, é que quando o tamanho amostral é próximo de 30, os testes parecem convergir em termos de decisão, ou seja, aprestaram as mesmas conclusões.

Tabela 13 – Resultados para os testes de adequação da distribuição Watson complexa utilizando os testes TRV, TKS, TCM e TAD.

Dados		n	k	$\hat{\kappa}$	Valor- p			
					TRV	TKS	TCM	TAD
Gorila	Macho	29	8	2406	0,0000	0,0000	0,0000	0,0000
	Fêmea	30	8	3140	0,0000	0,0200	0,0130	0,0000
Chimpanzé	Macho	28	8	2416	0,0060	0,0120	0,0000	0,0000
	Fêmea	26	8	2308	0,0000	0,0000	0,0000	0,0000
Orangotango	Macho	30	8	1702	0,0000	0,0130	0,0030	0,0010
	Fêmea	24	8	1937	0,6570	0,9680	0,9680	0,9660
Vértebra	Grande	23	6	990	0,0600	0,0960	0,0420	0,0190
	Pequena	23	6	976	0,0000	0,0000	0,0000	0,0000
Dígito 3		30	13	146	0,0000	0,0000	0,0000	0,0000
Conchas		21	3	209	0,0280	0,3150	0,1700	0,0380

Na Tabela 14 estão os resultados para os dados do crescimento em 18 ratos com 8 marcos ao longo de 8 períodos. Note que as estimativas de κ são valores altos e que houve uma divergência entre os testes no 30^o e 40^o dia. Entretanto, para os outros dias de crescimento ambos os testes não rejeitam a hipótese de que os dados provêm de uma distribuição Watson aos níveis de 1% e 5%.

Tabela 14 – Resultados para os testes de adequação para os dados dos crânios de $n = 18$ ratos e $k = 8$ marcos através de 8 momentos de crescimento utilizando os testes TRV, TKS, TCM e TAD.

Dias	$\hat{\kappa}$	Valor- p			
		TRV	TKS	TCM	TAD
7	5640	0,1990	0,3750	0,3210	0,3510
14	7789	0,0740	0,1340	0,0740	0,0470
21	10397	0,1660	0,1920	0,2510	0,1090
30	8776	0,0160	0,2520	0,2450	0,0450
40	12301	0,0420	0,0100	0,0170	0,0080
60	10614	0,8120	0,6770	0,8690	0,9220
90	10861	0,5480	0,8300	0,9500	0,7500
150	9679	0,9880	0,7530	0,5920	0,5970

4.9 CONCLUSÃO

Neste Capítulo foi introduzida uma aproximação da distribuição Watson complexa para a distribuição normal multivariada complexa singular. Essa aproximação foi utilizada em uma forma quadrática que resulta na distribuição Qui-quadrado, que foi utilizada como uma distribuição de referência para avaliar a bondade de ajuste da distribuição Watson complexa. Essa aproximação foi avaliada para diferentes configurações dos parâmetros da distribuição Watson complexa. Viu-se ainda, para quais valores de κ essa aproximação funcionava, já que a forma média não influencia diretamente na aproximação. A avaliação da aproximação foi feita através dos testes TRV, TKS, TCM e TAD, no qual concluiu-se que a aproximação é boa para valores de $\kappa \geq 10$.

Essa mesma aproximação foi avaliada com os valores estimados dos parâmetros da distribuição Watson complexa. Porém, os resultados não foram satisfatórios, pois foi observado que os tamanhos dos testes ficavam muito abaixo do valor teórico. Além disso, notou-se que a estimativa $\hat{\kappa}$ era a responsável pela falha dos testes. Assim, foi proposta uma transformação, na qual não era mais preciso estimar o valor de κ . Assim sendo, foram utilizados os testes TRV, TKS, TCM e TAD com o método *bootstrap*, em que uma distribuição empírica foi gerada para cada teste e os testes foram realizados normalmente. Concluiu-se que os testes utilizados com o método *bootstrap* alcançaram um bom desempenho em termos de tamanho e poder do teste. Por fim, foram utilizados vários bancos de dados, nos quais alguns deles se ajustaram a distribuição Watson complexa. Ademais, viu-se que quando o tamanho da amostra é pequena, os testes podem divergir em relação a rejeição da hipótese nula. Entretanto, quando o tamanho da amostra cresce os testes indicam concluir a mesma decisão.

5 BONDADDE DE AJUSTE PARA A DISTRIBUIÇÃO BINGHAM COMPLEXA

5.1 INTRODUÇÃO

A distribuição Bingham complexa, assim como a Watson complexa, é de grande importância na análise de forma. Essa distribuição foi proposta por Kent (1994), em que nesse trabalho são mostradas algumas propriedades como: relação com a Bingham real, estimação via MV e função geradora de momentos. Foram propostos três geradores para a Bingham complexa de números pseudo-aleatórios por Kent, Constable e Er (2004), no qual esses geradores foram comparados e conclui-se que os métodos de simulação possuem desempenho diferentes quando trabalhados em alta ou baixa concentração. Contudo, existe uma particularidade entre esses simuladores que diz respeito a rotação. A matriz de parâmetros é transformada na matriz de autovalores tornando a rotação fixa. Dessa forma, sempre são geradas amostras com a mesma forma média. À vista disso, é proposto um novo gerador para a distribuição Bingham complexa, no qual o pesquisador poderá especificar toda a matriz de parâmetros ou apenas os autovalores e a forma média, ocasionando em um controle maior na simulação de observações dessa distribuição.

Atualmente, não existem testes de bondade de ajuste para distribuição Bingham complexa. Pensando nisso, foi estudado o uso de uma aproximação da distribuição Bingham complexa para a distribuição normal multivariada complexa singular e uma transformação que recai em uma distribuição Qui-quadrado (da mesma forma como visto no Capítulo 4 sobre a distribuição Watson).

Será apresentada a distribuição Bingham complexa com suas principais propriedades na Seção 5.2; na Seção 5.3 é mostrado um simulador para a distribuição Bingham complexa; na Seção 5.4 é descrito a ortogonalização de Gram-Schmidt e o seu uso no simulador; na Seção 5.5 é mostrado uma aproximação da distribuição Bingham complexa para a normal multivariada complexa singular e uma transformação que leva na distribuição de Dirichlet; por fim, nas Seções 5.6 e 5.7 são mostrados resultados numéricos através de experimentos de simulação e aplicações em dados reais, respectivamente.

5.2 DISTRIBUIÇÃO BINGHAM COMPLEXA

A distribuição Bingham complexa é muito utilizada e é uma importante distribuição na esfera complexa para a análise de pré-formas. Uma variável aleatória \mathbf{Z} possui distribuição Bingham complexa, denotada por $\mathbb{CB}_{k-2}(\mathbf{A})$, se possuir a densidade

$$f(\mathbf{z}) = c(\mathbf{A})^{-1} \exp(\mathbf{z}^* \mathbf{A} \mathbf{z}),$$

na qual \mathbf{A} é uma matriz Hermitiana e $c(\mathbf{A})$ é a constante de normalização dada por

$$c(\mathbf{A}) = 2\pi^{k-1} \sum_{i=1}^{k-1} a_i \exp(\tilde{\lambda}_i), \quad a_i^{-1} = \prod_{i \neq j} (\tilde{\lambda}_j - \tilde{\lambda}_i),$$

com $\tilde{\lambda}_1 < \tilde{\lambda}_2 < \dots < \tilde{\lambda}_{k-1} = 0$ que são os autovalores da matriz \mathbf{A} .

Seja uma amostra aleatória $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ com distribuição $\mathbb{CB}_{k-2}(\mathbf{A})$. Considere a decomposição espectral (Teorema 1) da matriz $\mathbf{A} = \sum_{i=1}^{k-1} \tilde{\lambda}_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^*$ e da matriz $\mathbf{S} = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^* = \sum_{i=1}^{k-1} \lambda_i \mathbf{g}_i \mathbf{g}_i^*$. Dessa forma, tem-se que os estimadores de MV são dados por

$$\hat{\boldsymbol{\gamma}}_i = \mathbf{g}_i, \quad i = 1, 2, \dots, k-1,$$

e supondo alta concentração

$$\hat{\tilde{\lambda}}_i = -\frac{n}{\lambda_i}, \quad i = 1, 2, \dots, k-2.$$

5.3 SIMULADOR DA DISTRIBUIÇÃO BINGHAM COMPLEXA

A distribuição Bingham complexa possui três geradores propostos por Kent, Constante e Er (2004). Entretanto, a forma média é fixa e para gerar dados com uma forma média diferente é necessário rotacioná-los. Portanto, é proposto um gerador utilizando o método da aceitação-rejeição para a distribuição Bingham complexa em que o usuário poderá fornecer a forma média desejada, ou ainda, a própria matriz \mathbf{A} . Para encontrar um gerador, também será utilizado a desigualdade da Equação (4.2) proposta por Kent, Ganeiber e Mardia (2013), que é dada por

$$e^{-u} \leq e^{-(q-b)/2} \left(\frac{q/b}{1+2u/b} \right)^{q/2},$$

com $u > 0$, $q > 0$, $0 < b < q$ e essa desigualdade deve ser minimizada em relação a b . Fazendo $u = -\mathbf{z}^* \mathbf{A} \mathbf{z}$. Então, pela desigualdade temos que

$$\begin{aligned}
e^{-u} &\leq e^{-(q-b)/2} \left(\frac{q/b}{1+2u/b} \right)^{q/2}, \\
&\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} \left(\frac{1}{\mathbf{z}^* [\mathbf{I}_{k-1} - 2\mathbf{A}/b] \mathbf{z}} \right)^{q/2}, \\
&\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} \left(\frac{1}{\mathbf{z}^* \mathbf{B}^{-1} \mathbf{z}} \right)^{q/2}, \quad \mathbf{B}^{-1} = \mathbf{I}_{k-1} - 2\mathbf{A}/b.
\end{aligned}$$

Assim,

$$\begin{aligned}
\frac{c(\mathbf{A})}{c(\mathbf{A})} e^{-\mathbf{z}^* (-\mathbf{A}) \mathbf{z}} &\leq e^{-(q-b)/2} q^{q/2} b^{-q/2} (\mathbf{z}^* \mathbf{B}^{-1} \mathbf{z})^{-q/2} \frac{|\mathbf{B}|^{-1} \Gamma(q/2)}{|\mathbf{B}|^{-1} 2\pi^{q/2}} \frac{2\pi^{q/2}}{\Gamma(q/2)}, \\
f(\mathbf{z}) &\leq c(\mathbf{A})^{-1} e^{-(q-b)/2} q^{q/2} b^{-q/2} \frac{2\pi^{q/2}}{\Gamma(q/2)} |\mathbf{B}| g(\mathbf{z}),
\end{aligned}$$

em que $f(\mathbf{z})$ possui distribuição Bingham complexa e $g(\mathbf{z})$ com distribuição CGAC (4.3) fazendo $q = 2(k-1)$. Desta maneira, temos

$$M = c(\mathbf{A})^{-1} e^{-(q-b)/2} q^{q/2} b^{-q/2} \frac{2\pi^{q/2}}{\Gamma(q/2)} |\mathbf{B}|. \quad (5.1)$$

A equação (5.1) é uma função não-linear em relação a b e deverá ser minimizada. O Algoritmo 16 resume os passos para a geração de n amostras do método da aceitação-rejeição para a distribuição Bingham complexa. Com o intento de estudar a taxa de aceitação desse gerador, foi utilizado o seguinte esquema de simulação:

- Consideramos as dimensões $k = 3, 4, 5, 6, 7, 8, 9, 10$;
- A rotação é fixada em $\mathbf{U} = \mathbf{I}_{k-1}$;
- Para cada dimensão k , seja $\tilde{\boldsymbol{\lambda}}_{k-2}^j = \boldsymbol{\sigma}_j(k-1, k-2, \dots, 0)^\top$ o vetor de autovalores, no qual $\boldsymbol{\sigma}_j$ é o j -ésimo valor do vetor $\boldsymbol{\sigma} = (1, 5, 8, 10, 20, 30, 40, 50, 100, 500, 1000, 3000, 5000)$ com $j = 1, 2, \dots, 13$;
- Por fim, formando as matrizes $\mathbf{A}_{k-2}^j = \mathbf{U} \text{diag}(\tilde{\boldsymbol{\lambda}}_{k-2}^j) \mathbf{U}^*$.

Na Tabela 15 estão as taxas de aceitação para as configurações da dimensão do vetor de pré-forma $k-1$ e para vários valores do vetor de autovalores da matriz \mathbf{A} utilizando o Algoritmo 16. Pode-se ver que as taxas de aceitação são elevadas (acima de 90%) quando o valor de $\boldsymbol{\sigma}$ é pequeno (baixa concentração) considerando todas as dimensões. Em contra partida, quando a dimensão do vetor de pré-forma aumenta junto com o valor de $\boldsymbol{\sigma}$ (alta concentração), as probabilidades de aceitação diminuem.

Talvez seja difícil criar uma matriz Hermitiana para simular dados da Bingham complexa. O problema se torna mais difícil à medida que a dimensão do vetor de pré-forma aumenta. Em vista disso, é proposto uma maneira de criar essa matriz, fornecendo o vetor de autovalores e a forma média desejada. Isso será feito por meio de um método bastante conhecido, chamado de ortogonalização de Gram-Schmidt.

Tabela 15 – Probabilidade de aceitação ($1/M$) utilizando a distribuição CGAC para gerar amostras da distribuição Bingham complexa, usando várias combinações de autovalores e da dimensão k , para uma rotação fixa.

σ	k							
	3	4	5	6	7	8	9	10
1	0,9808	0,9739	0,9703	0,9682	0,9668	0,9657	0,9649	0,9643
3	0,8914	0,8434	0,8156	0,7972	0,7840	0,7741	0,7664	0,7602
5	0,8223	0,7406	0,6909	0,6566	0,6310	0,6111	0,5950	0,5816
8	0,7693	0,6643	0,6004	0,5561	0,5229	0,4967	0,4754	0,4575
10	0,7508	0,6388	0,5710	0,5241	0,4890	0,4615	0,4391	0,4203
20	0,7144	0,5907	0,5172	0,4671	0,4299	0,4010	0,3776	0,3581
30	0,7026	0,5757	0,5010	0,4502	0,4128	0,3837	0,3602	0,3408
40	0,6968	0,5684	0,4931	0,4421	0,4046	0,3755	0,3521	0,3327
50	0,6933	0,5641	0,4885	0,4374	0,3999	0,3708	0,3473	0,3280
100	0,6864	0,5556	0,4795	0,4282	0,3906	0,3615	0,3382	0,3189
1000	0,6803	0,5482	0,4716	0,4202	0,3826	0,3536	0,3303	0,3111
5000	0,6797	0,5475	0,4709	0,4195	0,3819	0,3529	0,3296	0,3104

Algoritmo 16: Simulador da distribuição Bingham complexa utilizando o método da aceitação-rejeição.

Entrada: Número de amostras n e a matriz de parâmetros \mathbf{A} .

Saída: n amostras da distribuição Bingham complexa.

1: Sejam $f(\mathbf{z})$ e $g(\mathbf{z})$ as densidades da distribuição Bingham complexa e CGAC, respectivamente.

2: Gere de forma independente uma amostra u de $U \sim \mathcal{U}(0, 1)$ e \mathbf{z} de $\mathbf{Z} \sim g$.

3: Aceite \mathbf{z} como sendo um amostra da Bingham complexa se

$$u \leq \frac{f(\mathbf{z})}{Mg(\mathbf{z})},$$

em que M é dado por (5.1) e deve ser minimizado em relação a b . Caso contrário, volte ao passo 2.

4: Repita os passos 2 e 3 até que se tenha n amostras.

5: Retorne as n amostras e a probabilidade de aceitação $1/M$.

5.4 ORTOGONALIZAÇÃO DE GRAM-SCHMIDT

O processo de ortogonalização de Gram-Schmidt é um método utilizado para obter uma base ortogonal a partir de uma base qualquer. Em outras palavras, o método consiste em transformar um conjunto de vetores linearmente independentes em um conjunto ortogonal de vetores, gerando o mesmo espaço vetorial (GENTLE, 1998, p. 102). A aplicação do processo de Gram-Schmidt nas colunas (vetores) de uma matriz \mathbf{C} produz o seguinte resultado

$$\mathbf{C} = \mathbf{QR},$$

em que \mathbf{Q} é uma matriz ortogonal e \mathbf{R} uma matriz triangular. Dessa forma, vamos considerar um conjunto de vetores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ que são linearmente independentes. Assim, o processo de Gram-Schmidt é dado por

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{x}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \\ \mathbf{u}_2 &= \mathbf{x}_2 - (\mathbf{x}_2^\top \mathbf{e}_1) \mathbf{e}_1, & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}, \\ & \vdots & & \\ \mathbf{u}_n &= \mathbf{x}_n - \sum_{j=1}^{n-1} (\mathbf{x}_n^\top \mathbf{e}_j) \mathbf{e}_j, & \mathbf{e}_n &= \frac{\mathbf{u}_n}{\|\mathbf{u}_n\|}. \end{aligned}$$

Portanto, temos a matriz ortogonal $\mathbf{Q} = (\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_n)$ (BOYD; VANDENBERGHE, 2018, p. 97-99,190) e a matriz

$$\mathbf{R} = \begin{pmatrix} \langle \mathbf{x}_1, \mathbf{e}_1 \rangle & \langle \mathbf{x}_2, \mathbf{e}_1 \rangle & \langle \mathbf{x}_3, \mathbf{e}_1 \rangle & \dots & \langle \mathbf{x}_n, \mathbf{e}_1 \rangle \\ 0 & \langle \mathbf{x}_2, \mathbf{e}_2 \rangle & \langle \mathbf{x}_2, \mathbf{e}_3 \rangle & \dots & \langle \mathbf{x}_n, \mathbf{e}_2 \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \langle \mathbf{x}_n, \mathbf{e}_n \rangle \end{pmatrix}.$$

Essa decomposição também pode ser utilizada em vetores complexos, basta utilizar o transposto conjugado onde existir o operador transposto $(\cdot)^\top$ (HUANG; TSAI, 2011). Esse processo será utilizado para construir uma matriz ortogonal para ser utilizada no gerador da Bingham complexa. No qual, será fornecido a forma média desejada e o processo encontrará a matriz ortogonal, mantendo a forma média que foi estabelecida.

5.4.1 Simulador da Distribuição Bingham Complexa via Gram-Schmidt

Como ilustração, considere matriz de configuração da equação (4.5) utilizada no simulador da distribuição Watson complexa. Após a utilização dos precedimentos de remoção da locação e escala, tem-se o vetor

$$\boldsymbol{\mu} = \frac{\boldsymbol{\omega}_{\mu_2}}{\|\boldsymbol{\omega}_{\mu_2}\|} = \begin{pmatrix} -0,4629 - 0,3086i \\ 0,4629 - 0,3086i \\ 0,0000 + 0,6172i \end{pmatrix}.$$

Perceba que para utilizar o processo de Gram-Schmidt é necessário fornecer uma matriz. À vista disso, considerando que $k - 1 = 3$ (mantendo a notação da pré-forma), uma alternativa é preencher a matriz com uma matriz identidade sem a primeira coluna, ou seja,

$$\mathbf{C} = [\boldsymbol{\mu} | \mathbf{I}_{-1}],$$

no qual \mathbf{I}_{-1} é a matriz identidade de dimensão 3 sem a primeira coluna. Agora, basta utilizar o processo de ortogonalização e veja que a primeira operação do processo não altera a forma média. Assim, após o processo temos que a matriz $\mathbf{C} = \mathbf{QR}$, no qual

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} -0,4629 - 0,3086i & 0,1433 + 0,3438i & 0,4120 - 0,618i \\ 0,4629 - 0,3086i & 0,8309 + 0,0000i & 0,0000 + 0,000i \\ 0,0000 + 0,6172i & 0,2292 - 0,3438i & 0,6695 + 0,000i \end{bmatrix} \\ &= [\boldsymbol{\mu} | \mathbf{Q}_{-1}], \end{aligned}$$

em que \mathbf{Q}_{-1} é a matriz \mathbf{Q} sem a primeira coluna. Para criar a matriz de parâmetros da Bingham complexa, considere a decomposição espectral da matriz \mathbf{A} ,

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*.$$

Agora, faça $\mathbf{U} = [\mathbf{Q}_{-1} | \boldsymbol{\mu}]$ e $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, 0)$. Assim, basta informa os autovalores e teremos a matriz de parâmetros desejada. Como ilustração, veja a Figura 33, onde estão algumas simulações para $n = 50$ amostras da distribuição Bingham complexa considerado quatro pares de autovalores diferentes. Perceba que, à medida que os autovalores $\lambda_1, \lambda_2 \rightarrow -\infty$, os pontos vão ficando mais concentrados ao redor da forma média e isso é caracterizado como alta concentração.

No Algoritmo 17 estão descritos os passos para obter uma amostra de um variável aleatória Bingham fornecendo a forma média e a matriz de autovalores.

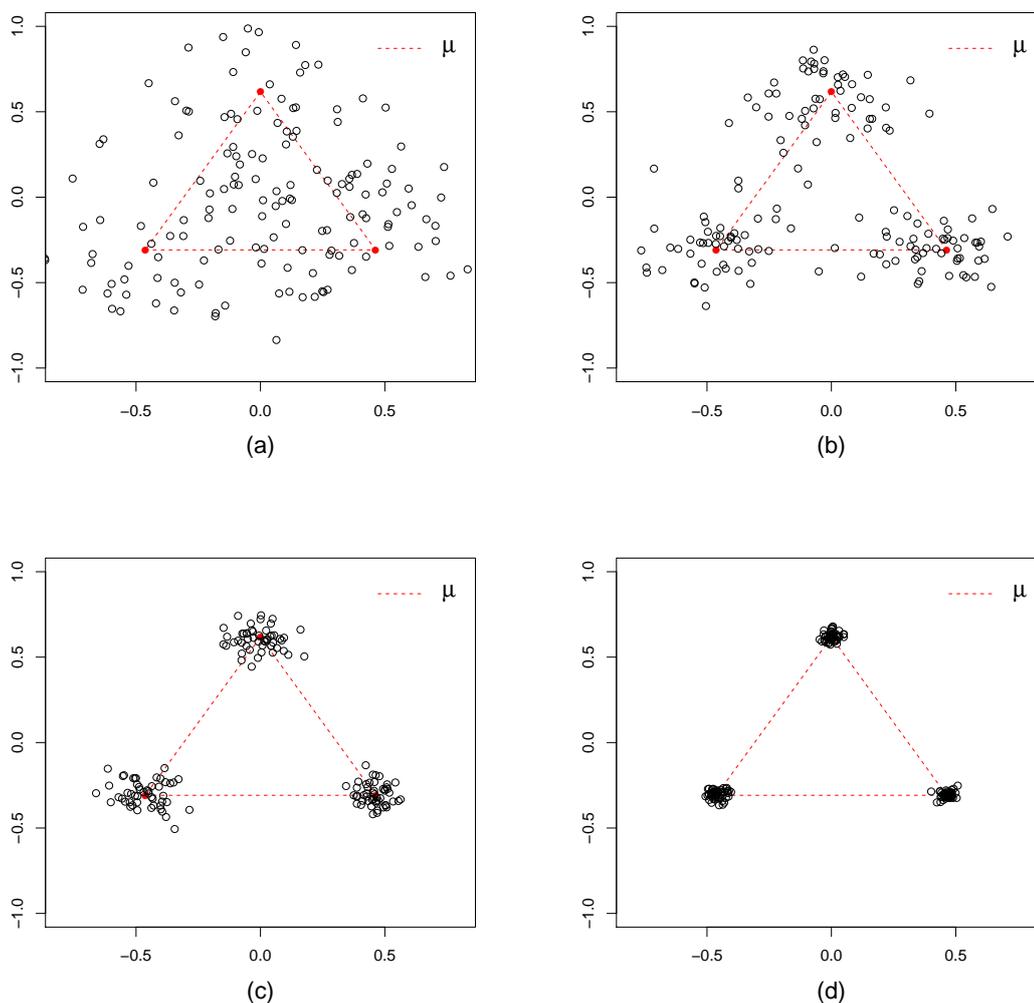


Figura 33 – Amostras da distribuição Bingham complexa, considerando a forma média μ da equação (4.9), $n = 50$ e diferentes pares de autovalores: (a) $\lambda_1 = -2, \lambda_2 = -1$; (b) $\lambda_1 = -20, \lambda_2 = -10$; (c) $\lambda_1 = -100, \lambda_2 = -50$; (d) $\lambda_1 = -1000, \lambda_2 = -500$.

Algoritmo 17: Simulador da Bingham complexa utilizando o método da aceitação-rejeição por Gram-Schmidt.

Entrada: Número de amostras n , forma média μ e a matriz de autovalores Λ .

Saída: n amostras da distribuição Bingham complexa.

1: Crie a matriz $\mathbf{C} = [\mu | \mathbf{I}_{-1}]$;

2: Utilize o processo de ortogonalização de Gram-Schmidt, resultando em $\mathbf{C} = \mathbf{QR}$;

3: Faça $\mathbf{U} = [\mathbf{Q}_{-1} | \mu]$;

4: Por fim, faça $\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^*$ e utilize o Algoritmo 16.

5.5 APROXIMAÇÃO DA DISTRIBUIÇÃO BINGHAM COMPLEXA PELA NORMAL MULTIVARIADA COMPLEXA SINGULAR

Inicialmente, considere uma variável aleatória $\mathbf{Z} \sim \mathbb{C}B_{k-2}(\mathbf{A})$. Assim, constante de normalização da distribuição Bingham complexa, que é dada por

$$c(\mathbf{A}) = 2\pi^{k-1} \sum_{j=1}^{k-1} a_j \exp(\tilde{\lambda}_j), \quad a_j^{-1} = \prod_{i \neq j} (\tilde{\lambda}_j - \tilde{\lambda}_i),$$

em que $\tilde{\lambda}_1 < \tilde{\lambda}_2 < \dots < \tilde{\lambda}_{k-1} = 0$ são os autovalores da matriz \mathbf{A} . Observe que

$$c(\mathbf{A}) = 2\pi^{k-1} \left[\frac{\exp(\tilde{\lambda}_1)}{\prod_{i \neq 1} (\tilde{\lambda}_1 - \tilde{\lambda}_i)} + \frac{\exp(\tilde{\lambda}_2)}{\prod_{i \neq 2} (\tilde{\lambda}_2 - \tilde{\lambda}_i)} + \dots + \frac{\exp(\tilde{\lambda}_{k-2})}{\prod_{i \neq k-2} (\tilde{\lambda}_{k-2} - \tilde{\lambda}_i)} + \frac{1}{\prod_{i=1}^{k-2} (-\tilde{\lambda}_i)} \right].$$

Veja que

$$\frac{\exp(\tilde{\lambda}_r)}{\prod_{i \neq r} (\tilde{\lambda}_r - \tilde{\lambda}_i)}, \quad r = 1, 2, \dots, k-2,$$

o numerador é dado por uma função com decaimento exponencial e o denominador é um polinômio de grau $k-3$ em $\tilde{\lambda}_r$. Assim sendo, supondo alta concentração, dispomos que

$$\lim_{\tilde{\lambda}_r \rightarrow -\infty} \frac{\exp(\tilde{\lambda}_r)}{\prod_{i \neq r} (\tilde{\lambda}_r - \tilde{\lambda}_i)} = 0, \quad \text{para } r = 1, 2, \dots, k-2.$$

Em consequência disso, a constante de normalização é recalculada, tornando-se

$$\begin{aligned} c(\mathbf{A}) &= 2\pi^{k-1} \left(\prod_{i=1}^{k-2} -\tilde{\lambda}_i \right)^{-1} \\ &= 2^{1+k-2+2-k} \pi^{k-1} \left(\prod_{i=1}^{k-2} -\tilde{\lambda}_i \right)^{-1} \\ &= (2\pi)^{k-1} 2^{-(k-2)} \left(\prod_{i=1}^{k-2} -\tilde{\lambda}_i \right)^{-1} \\ &= (2\pi)^{k-1} \prod_{i=1}^{k-2} (-2\tilde{\lambda}_i)^{-1}. \end{aligned}$$

A densidade da Bingham complexa, supondo alta concentração, é aproximadamente (KENT, 1994)

$$f(\mathbf{z}) = \frac{\exp\{\mathbf{z}^* \mathbf{A} \mathbf{z}\}}{2\pi^{k-1} \left[\prod_{i=1}^{k-2} (-\tilde{\lambda}_i) \right]^{-1}} = \frac{\exp\left\{-\frac{1}{2} \mathbf{z}^* (-2\mathbf{A}) \mathbf{z}\right\}}{(2\pi)^{k-1} \prod_{i=1}^{k-2} (-2\tilde{\lambda}_i)^{-1}}.$$

Perceba que o posto de \mathbf{A} é igual a $k-2$, pois \mathbf{A} possui $k-2$ autovalores diferentes de zero. Portanto, pode-se ver que $\mathbf{Z} \approx \mathbb{C}SN_{k-1}(\mathbf{0}, -\frac{1}{2}\mathbf{A}^-)$. Utilizando o resultado da equação (4.15), tem-se que

$$-2\mathbf{Z}^*\mathbf{A}\mathbf{Z} \approx \chi_{2(k-2)}^2. \quad (5.2)$$

Esse resultado será utilizado na criação de testes de bondade de ajuste para a distribuição Bingham complexa. Perceba que distribuição independe dos valores da matriz \mathbf{A} , depende apenas do posto da matriz que é $k - 2$. Ademais, será feita uma análise para verificar a qualidade dessa aproximação para diversas configurações dos parâmetros de concentração. Assim, identificando para quais valores da concentração, a aproximação em (5.2) é satisfeita.

5.5.1 Avaliação da Aproximação da Distribuição Qui-quadrado - Bingham Complexa

Foram geradas $n = 200$ amostras da distribuição $\mathbb{C}B_{6-2}(\mathbf{A})$, considerando a matriz

$$\mathbf{A} = \begin{bmatrix} -400 & 0 & 0 & 0 & 0 \\ 0 & -300 & 0 & 0 & 0 \\ 0 & 0 & -200 & 0 & 0 \\ 0 & 0 & 0 & -100 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Utilizando a aproximação da equação (5.2), obtém-se uma distribuição χ_8^2 . Pela Figura 34, percebe-se que a distribuição χ_8^2 se ajusta perfeitamente utilizando o resultado em (5.2).

Com o propósito de ampliar esse resultado, serão utilizadas 5000 réplicas de Monte Carlo para averiguar o tamanho do TRV (4.17), TKS (4.19), TCM (4.20) e TAD (4.21) para diferentes configurações da matriz \mathbf{A} , número de amostras e utilizando a dimensão do vetor de pré-forma $k - 1 = 5$. As configurações utilizadas foram:

- $\mathbf{A} = \sigma \times \text{diag}(-4, -3, -2, -1, 0)$;
- $\sigma = 1, 3, 5, 8, 10, 20, 30, 40, 50, 100, 1000, 5000$;
- $n = 20, 30, 40, 50, 100, 200$.

Na Tabela 16, estão os resultados da simulação utilizando um tamanho de $\alpha = 5\%$. Nota-se que para os valores de $\sigma \geq 8$, as taxas de rejeição se aproximam do nível de 5%, para todos os valores de n . Portanto, conclui-se que os testes funcionam para essa aproximação. No Apêndice B, nas Tabelas 49 e 50 estão as taxas de rejeição para a simulação utilizando $\alpha = 1\%$ e 10%, respectivamente. A mesma situação ocorre, indicando que os testes também funcionaram para $\sigma \geq 8$.

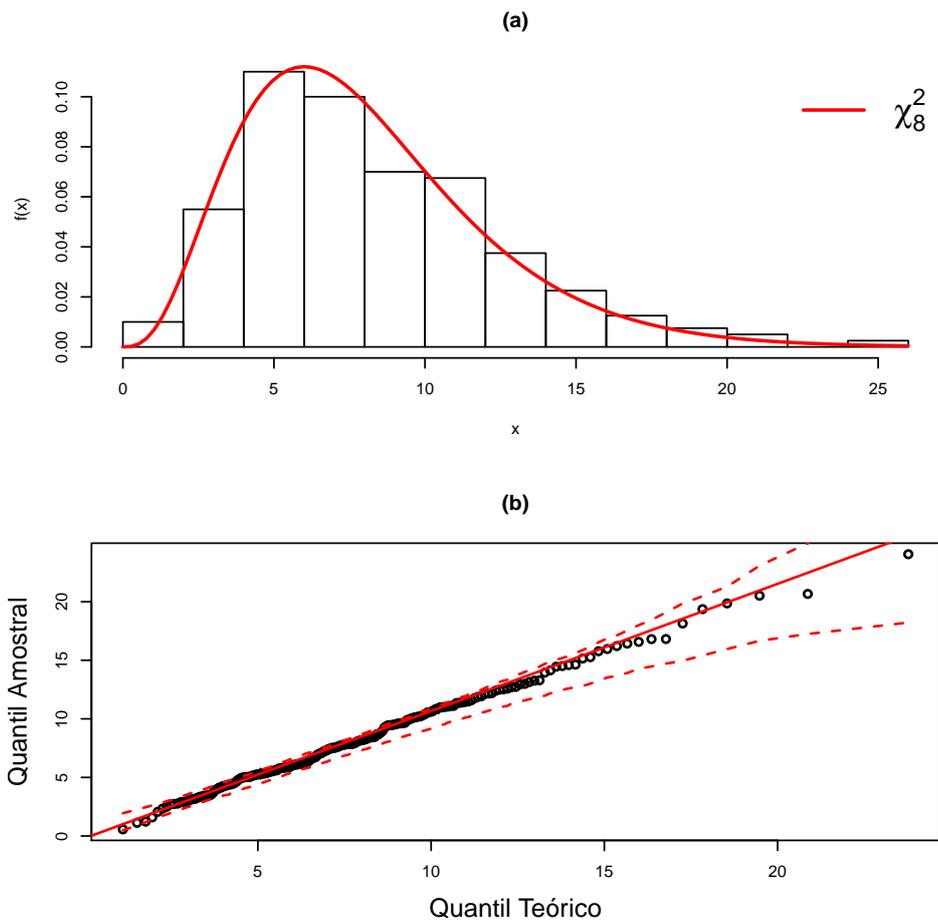


Figura 34 – Análise gráfica utilizando a aproximação (5.2), utilizando dados da distribuição Bingham complexa com $k = 6$, $\mathbf{A} = \text{diag}(-400, -300, -200, -100, 0)$ e $n = 200$. Gráfico (a): Histograma; (b): Quantil-Quantil.

Tabela 16 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível de significância de 5%.

Teste	n	σ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	0,1592	0,0450	0,0484	0,0528	0,0480	0,0486	0,0504	0,0488	0,0478	0,0494	0,0500
	30	1,0000	0,2530	0,0490	0,0462	0,0476	0,0486	0,0438	0,0464	0,0496	0,0508	0,0506	0,0516
	40	1,0000	0,3530	0,0482	0,0480	0,0512	0,0540	0,0508	0,0502	0,0540	0,0524	0,0504	0,0484
	50	1,0000	0,4318	0,0580	0,0492	0,0562	0,0506	0,0472	0,0438	0,0462	0,0488	0,0464	0,0570
	100	1,0000	0,7620	0,0682	0,0498	0,0496	0,0484	0,0496	0,0540	0,0486	0,0496	0,0506	0,0478
	200	1,0000	0,9774	0,0890	0,0486	0,0528	0,0490	0,0456	0,0534	0,0502	0,0472	0,0480	0,0522
TKS	20	1,0000	0,1774	0,0502	0,0506	0,0464	0,0496	0,0498	0,0500	0,0444	0,0530	0,0494	0,0446
	30	1,0000	0,2936	0,0544	0,0476	0,0494	0,0436	0,0478	0,0516	0,0514	0,0506	0,0480	0,0536
	40	1,0000	0,3854	0,0510	0,0486	0,0488	0,0516	0,0460	0,0504	0,0526	0,0478	0,0542	0,0540
	50	1,0000	0,4782	0,0506	0,0510	0,0496	0,0506	0,0524	0,0452	0,0464	0,0478	0,0454	0,0506
	100	1,0000	0,8342	0,0546	0,0468	0,0456	0,0434	0,0432	0,0460	0,0410	0,0416	0,0418	0,0416
	200	1,0000	0,9960	0,0668	0,0462	0,0516	0,0440	0,0472	0,0446	0,0468	0,0496	0,0454	0,0436
TCM	20	1,0000	0,2050	0,0524	0,0514	0,0484	0,0512	0,0494	0,0510	0,0472	0,0482	0,0490	0,0458
	30	1,0000	0,3346	0,0540	0,0472	0,0466	0,0476	0,0464	0,0490	0,0548	0,0504	0,0498	0,0520
	40	1,0000	0,4322	0,0528	0,0496	0,0522	0,0514	0,0502	0,0508	0,0538	0,0494	0,0516	0,0506
	50	1,0000	0,5380	0,0578	0,0482	0,0528	0,0500	0,0520	0,0464	0,0434	0,0506	0,0474	0,0534
	100	1,0000	0,8744	0,0646	0,0506	0,0500	0,0498	0,0490	0,0526	0,0476	0,0462	0,0508	0,0502
	200	1,0000	0,9976	0,0768	0,0458	0,0558	0,0472	0,0482	0,0482	0,0494	0,0498	0,0498	0,0488
TAD	20	1,0000	0,1934	0,0508	0,0498	0,0474	0,0486	0,0498	0,0506	0,0462	0,0482	0,0488	0,0480
	30	1,0000	0,3312	0,0520	0,0470	0,0460	0,0494	0,0502	0,0470	0,0520	0,0510	0,0478	0,0514
	40	1,0000	0,4496	0,0480	0,0490	0,0512	0,0506	0,0502	0,0502	0,0558	0,0518	0,0496	0,0518
	50	1,0000	0,5666	0,0576	0,0502	0,0544	0,0520	0,0500	0,0468	0,0454	0,0518	0,0446	0,0544
	100	1,0000	0,9204	0,0674	0,0492	0,0486	0,0486	0,0484	0,0536	0,0474	0,0470	0,0504	0,0480
	200	1,0000	0,9994	0,0910	0,0492	0,0554	0,0458	0,0472	0,0506	0,0506	0,0498	0,0488	0,0476

Evidentemente, os parâmetros da distribuição Bingham complexa não são conhecidos. Dessa forma, a proposta é utilizar o resultado (5.2) com as estimativas dos parâmetros via MV, como feito anteriormente para a distribuição Watson complexa. Os testes foram utilizados com os parâmetros estimados, porém, assim como na distribuição Watson complexa, os resultados não foram satisfatórios. O método *bootstrap* foi utilizado nos testes, mesmo assim, também não produziram melhoras. Dessa forma, foi utilizado a mesma transformação proposta para os testes de adequação da distribuição Watson complexa, resultando em uma distribuição de Dirichlet.

Considere uma amostra aleatória $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ de $\mathbf{Z} \sim \mathbb{CB}_{k-2}(\mathbf{A})$. Assim, utilizando a aproximação em (5.2) e fazendo

$$V_i = -2\mathbf{Z}_i^* \mathbf{A}^{-1} \mathbf{Z}_i, \text{ para } i = 1, \dots, n,$$

temos que $V_i \approx \chi_{2k-4}^2$, equivalentemente, $V_i \approx \text{Gama}(k-2, \frac{1}{2})$. Assim sendo, considere a transformação

$$\begin{aligned} Y_j &= \frac{V_j}{\sum_{i=1}^n V_i} = \frac{-2\mathbf{Z}_j^* \mathbf{A}^{-1} \mathbf{Z}_j}{\sum_{i=1}^n -2\mathbf{Z}_i^* \mathbf{A}^{-1} \mathbf{Z}_i} \\ &= \frac{\mathbf{Z}_j^* \mathbf{A}^{-1} \mathbf{Z}_j}{\sum_{i=1}^n \mathbf{Z}_i^* \mathbf{A}^{-1} \mathbf{Z}_i}, \quad j = 1, \dots, n. \end{aligned} \quad (5.3)$$

Desse modo, pelo Teorema 5, Y_1, \dots, Y_{n-1} possui distribuição conjunta $Dr(a_1, \dots, a_n)$, no qual $a_i = k-2$ ($i = 1, \dots, n$). A densidade conjunta de (5.3) é uma distribuição simétrica de Dirichlet e possui a densidade

$$f(y_1, \dots, y_{n-1}) = \frac{\Gamma(na)}{\Gamma(a)^n} \left(\prod_{i=1}^{n-1} y_i^{a-1} \right) \left(1 - \sum_{i=1}^{n-1} y_i \right)^{a-1},$$

para $(y_1, \dots, y_{n-1}) \in \mathbb{S}$ e $a = k-2$. A distribuição marginal para cada Y_i ($i = 1, \dots, n-1$) é Beta com a densidade dada por

$$f(y_i) = \frac{\Gamma(na)}{\Gamma(a)\Gamma((n-1)a)} y_i^{a-1} (1-y_i)^{(n-1)a-1}, \quad y_i \in (0, 1). \quad (5.4)$$

Dessa forma, será utilizada a transformação (5.3) com a matriz estimada via MV e os testes serão utilizados para verificar se os dados provêm da distribuição Beta com densidade (5.4) juntamente com o método *bootstrap* para estimar a distribuição das estatísticas de testes empiricamente.

5.6 AVALIAÇÃO NUMÉRICA - SIMULAÇÃO

5.6.1 Tamanho do Teste

Para a simulação de tamanho, foram utilizadas as seguintes configurações:

- $\mathbf{A} = \sigma \times \text{diag}(-4, -3, -2, -1, 0)$;
- $\sigma = 8, 10, 20, 30, 40, 50, 1000, 5000$;
- $n = 20, 30, 40, 50, 100, 200$.

Na Tabela 17 estão os resultados das simulações considerando os valores de $\sigma = 8, 40, 5000$, utilizando $n = 20, 30, 40, 50, 100, 200$ e considerando os tamanhos $\alpha = 1\%, 5\%, 10\%$ para todos os testes. Analisando o TRV, percebe-se que para $n = 20, 30$ e 40 os valores dos tamanhos estimados ficaram um pouco acima dos verdadeiros valores de α . Para os testes TKS e TCM, em todas as configurações, os tamanhos estimados ficaram muito próximos dos valores verdadeiros. Por fim, o TAD também resultou em um bom desempenho, exceto quando $\sigma = 8$ para $n = 20$ e 30 , porém quando $\sigma \geq 10$ esse problema não ocorre (ver Apêndice B). De modo geral, quando o tamanho da amostra cresce, o valor da probabilidade de rejeição se aproxima ainda mais do verdadeiro nível nominal do teste. Para verificar outras configurações, ver as Tabelas 51, 52 e 53 no Apêndice B.

Tabela 17 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, considerando a matriz $\mathbf{A} = \sigma \times \text{diag}(-4, -3, -2, -1, 0)$, para $\sigma = 10, 50, 5000$, $n = 20, 30, 40, 50, 100, 200$ e os níveis de significância de $\alpha = 1\%, 5\%, 10\%$.

	σ	8			40			5000			
		α	1%	5%	10%	1%	5%	10%	1%	5%	10%
TRV	n	20	0,0182	0,0774	0,1398	0,0224	0,0892	0,1570	0,0176	0,0732	0,1356
		30	0,0146	0,0652	0,1218	0,0162	0,0710	0,1324	0,0160	0,0642	0,1134
		40	0,0122	0,0640	0,1204	0,0136	0,0630	0,1228	0,0124	0,0590	0,1128
		50	0,0116	0,0584	0,1080	0,0118	0,0616	0,1206	0,0126	0,0570	0,1158
		100	0,0102	0,0596	0,1128	0,0104	0,0542	0,1030	0,0082	0,0422	0,0920
		200	0,0112	0,0500	0,1052	0,0120	0,0572	0,1080	0,0102	0,0556	0,1096
TKS	n	20	0,0106	0,0508	0,1084	0,0094	0,0600	0,1122	0,0120	0,0570	0,1134
		30	0,0096	0,0490	0,0994	0,0092	0,0524	0,1018	0,0096	0,0512	0,1032
		40	0,0086	0,0454	0,0988	0,0106	0,0488	0,0998	0,0080	0,0468	0,0966
		50	0,0082	0,0492	0,1020	0,0094	0,0504	0,1008	0,0118	0,0504	0,0976
		100	0,0088	0,0474	0,0994	0,0094	0,0530	0,1008	0,0076	0,0468	0,0940
		200	0,0112	0,0518	0,1026	0,0102	0,0508	0,0988	0,0118	0,0516	0,1028
TCM	n	20	0,0114	0,0560	0,1020	0,0124	0,0622	0,1192	0,0114	0,0588	0,1146
		30	0,0104	0,0470	0,0970	0,0106	0,0558	0,1094	0,0112	0,0524	0,1074
		40	0,0090	0,0462	0,1012	0,0126	0,0522	0,0988	0,0086	0,0430	0,0932
		50	0,0086	0,0494	0,0996	0,0088	0,0508	0,1020	0,0138	0,0518	0,0988
		100	0,0086	0,0520	0,1054	0,0104	0,0542	0,1072	0,0078	0,0446	0,0930
		200	0,0098	0,0512	0,0982	0,0106	0,0520	0,0974	0,0100	0,0516	0,1052
TAD	n	20	0,0048	0,0376	0,0894	0,0050	0,0432	0,0958	0,0072	0,0452	0,0962
		30	0,0064	0,0374	0,0832	0,0070	0,0434	0,0988	0,0080	0,0462	0,0970
		40	0,0082	0,0418	0,0924	0,0092	0,0426	0,0882	0,0064	0,0392	0,0870
		50	0,0062	0,0434	0,0878	0,0070	0,0468	0,0944	0,0094	0,0478	0,0908
		100	0,0082	0,0528	0,1014	0,0102	0,0524	0,1010	0,0072	0,0422	0,0852
		200	0,0102	0,0500	0,1008	0,0102	0,0498	0,0996	0,0100	0,0512	0,1028

5.6.2 Poder do Teste

Para análise do poder foi utilizado o mesmo mecanismo de contaminação na amostra, como feito anteriormente para a distribuição Watson complexa. O mecanismo será utilizado com uma contaminação em 10% do tamanho amostral de outra variável com distribuição Bingham complexa com uma forma média diferente. O esquema de contaminação e avaliação do poder é descrito a seguir:

- $\mathbf{A} = \mathbf{U}[\sigma \times \text{diag}(-4, -3, -2, -1, 0)]\mathbf{U}^*$;
- Com os valores de $\sigma = 8, 10, 20, 30, 40, 50, 100, 1000, 5000$;
- A matriz de rotação encontrar-se-á fixada em $\mathbf{U} = \mathbf{I}$, em consequência, a forma média nesse caso é a última coluna na matriz identidade, isto é, $\boldsymbol{\mu} = (0, 0, 0, 0, 1)^\top$;
- Foram usados sete valores para a contaminação da forma média, os mesmos utilizados para a análise de poder da distribuição Watson complexa através do vetor $\boldsymbol{\varepsilon}$ dado em (4.25). Gerando as mesmas formas médias calculadas através da equação (4.26). Para geração dessas amostras, foi utilizado o simulador da distribuição Bingham complexa via

Gram-Schmidt no qual os passos desse simulador são descritos no Algoritmo 17.

Considerando $\alpha = 5\%$, concluiu-se que a partir de $\sigma = 20$ o poder começa a aumentar à medida que o valor da contaminação e o número de amostras aumentam, chegando em torno de 38% a 50% para $n = 200$ nos quatro testes (ver Tabela 65). Além disso, percebeu-se que para $\sigma = 100$ o poder se torna atrativo, chegando em altos valores a partir de ϵ_4 como mostrado na Tabela 18. De forma geral, para um dado tamanho de amostra n fixo, à medida que o valor da contaminação aumenta, o poder também aumenta.

Com o intuito de melhorar a visualização dos resultados da Tabela 18, observe na Figura 35 as curvas de poder para os testes de bondade de ajuste, considerando $n = 20, 40, 200$. Verifica-se que quanto maior o tamanho da amostra, maior o poder.

Tabela 18 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\sigma = 100$, $n = 20, 30, 40, 50, 100, 200$, nível de significância de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0666	0,0554	0,0478	0,0566	0,0826	0,1180	0,1492
	30	0,0580	0,0612	0,0738	0,1386	0,2118	0,2766	0,3472
	40	0,0580	0,0608	0,1150	0,2198	0,3408	0,4348	0,5254
	50	0,0596	0,0764	0,1590	0,3156	0,4578	0,5936	0,6908
	100	0,0556	0,1566	0,3856	0,6808	0,8524	0,9382	0,9740
	200	0,1036	0,2946	0,7232	0,9490	0,9942	0,9996	0,9998
TKS	20	0,0478	0,0456	0,0558	0,0796	0,1034	0,1420	0,1770
	30	0,0452	0,0584	0,0872	0,1476	0,2238	0,2844	0,3448
	40	0,0504	0,0642	0,1240	0,2146	0,3438	0,4214	0,4902
	50	0,0586	0,0734	0,1488	0,2950	0,4266	0,5460	0,6220
	100	0,0602	0,1414	0,3386	0,5968	0,7882	0,8980	0,9484
	200	0,0926	0,2334	0,6146	0,9044	0,9878	0,9980	0,9994
TCM	20	0,0506	0,0446	0,0562	0,0798	0,1122	0,1526	0,1938
	30	0,0484	0,0594	0,0878	0,1642	0,2562	0,3264	0,4058
	40	0,0512	0,0640	0,1332	0,2450	0,3900	0,5086	0,5840
	50	0,0626	0,0756	0,1724	0,3526	0,5092	0,6516	0,7362
	100	0,0578	0,1552	0,3916	0,6980	0,8812	0,9624	0,9872
	200	0,0962	0,2826	0,7216	0,9640	0,9980	0,9998	1,0000
TAD	20	0,0378	0,0412	0,0620	0,1020	0,1540	0,2006	0,2532
	30	0,0432	0,0630	0,1072	0,1936	0,3106	0,3972	0,4974
	40	0,0490	0,0700	0,1612	0,2970	0,4560	0,5796	0,6688
	50	0,0636	0,0900	0,2016	0,4078	0,5770	0,7234	0,8138
	100	0,0632	0,1734	0,4422	0,7554	0,9172	0,9832	0,9952
	200	0,1054	0,3142	0,7692	0,9768	0,9990	1,0000	1,0000

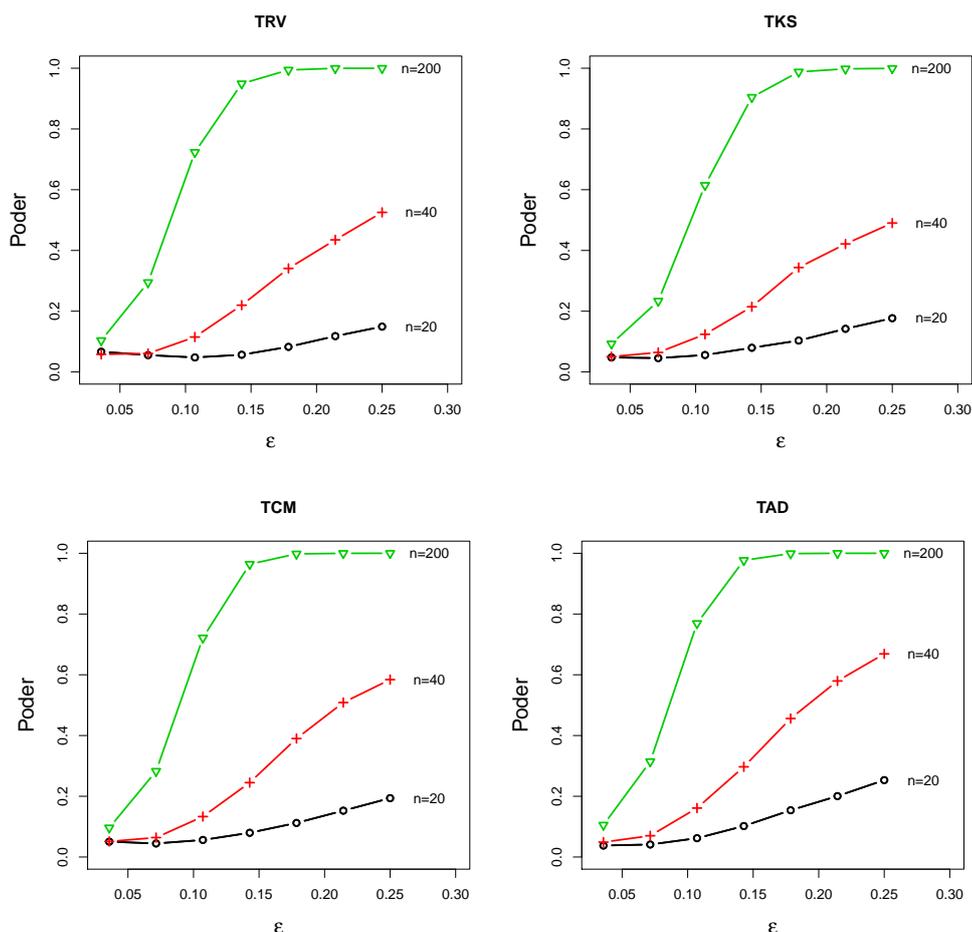


Figura 35 – Curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ϵ com 7 valores, $\sigma = 100$ e $n = 20, 40, 200$.

Na Figura 36 pode-se analisar o desempenho entre os testes em relação ao poder. Veja que, em todas as situações do número de amostras o TAD obteve um poder maior do que os demais testes. Entretanto, quando o tamanho da amostra cresce, a diferença entre os testes é pequena.

Considerando uma concentração um pouco maior com $\sigma = 1000$, veja na Figura 37 que o TAD obteve uma curva de poder bem acima dos outros testes quando $n = 20$ e, quando o número de amostras aumenta, rapidamente o poder do teste se aproxima de 1. As análises do poder também foram realizadas considerando $\alpha = 1\%$ e 10% , produzindo resultados similares. Para mais detalhes, ver Apêndice B.

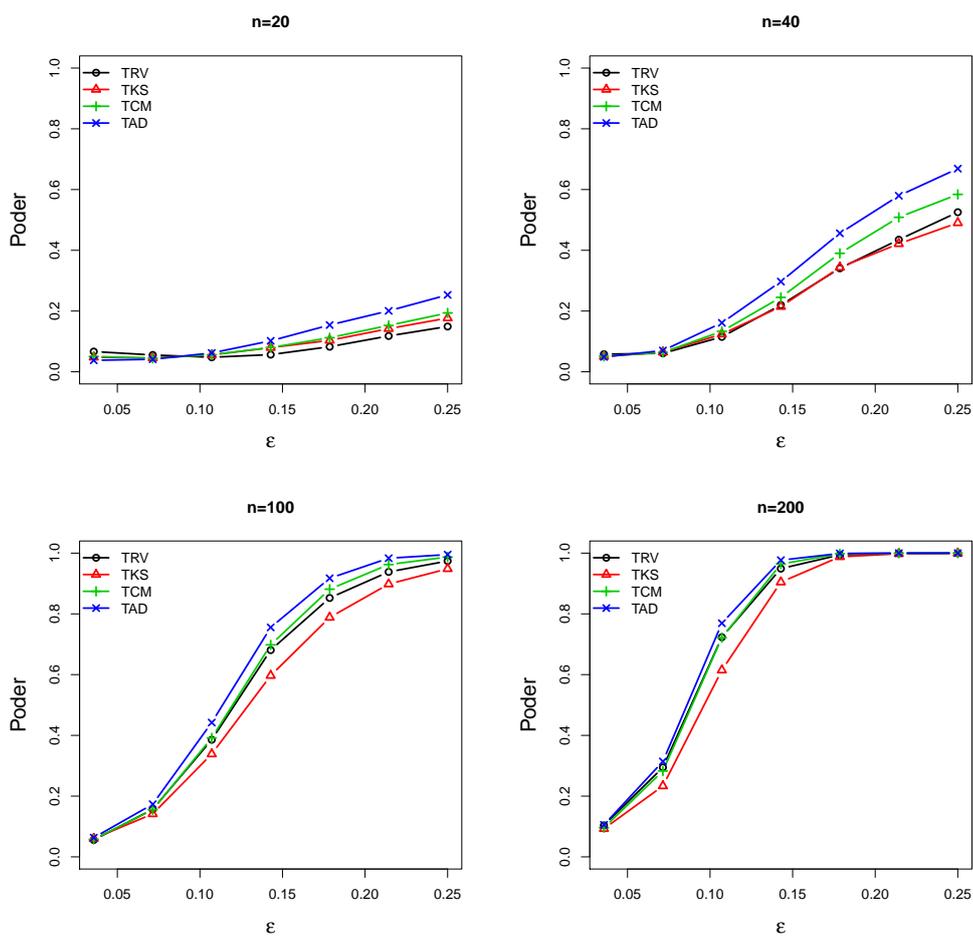


Figura 36 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ϵ com 7 valores, $\sigma = 100$ e $n = 20, 40, 200$.

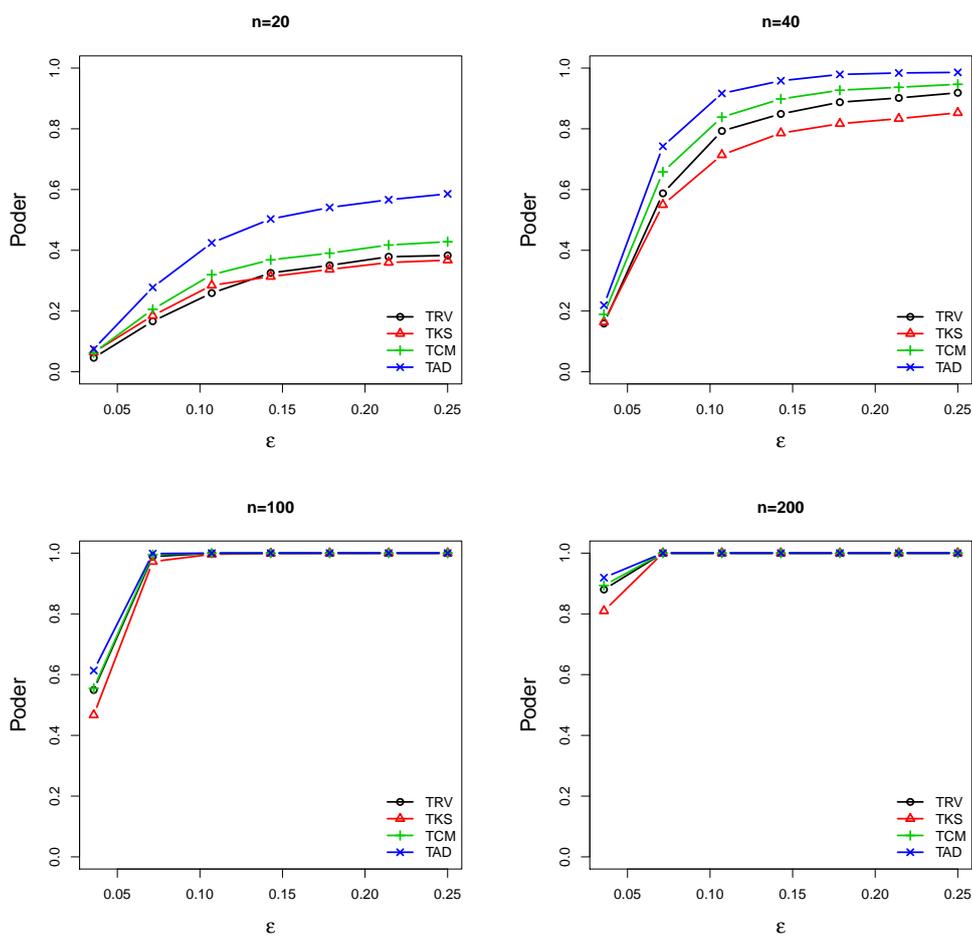


Figura 37 – Comparação entre as curvas de poder para os testes de bondade de ajuste, considerando o vetor de contaminação ϵ com 7 valores, $\sigma = 1000$ e $n = 20, 40, 200$.

5.7 APLICAÇÃO EM DADOS REAIS

Os dados utilizados são os mesmos discutidos na subseção 3.11 e na seção 4.8. A Tabela 19 possui um resumo de alguns valores para os testes de adequação, como: o tamanho da amostra n , número de marcos k , estimativas do primeiro e penúltimo autovalores da matriz \mathbf{A} e os valores de significância para cada teste em todos os bancos de dados. Note que todos autovalores estimados são grandes em valor absoluto e o menor deles é 22, para os dados referentes ao dígito 3. Portanto, se os dados possuírem distribuição Bingham complexa, a aproximação para a distribuição normal multivariada complexa singular será boa e, conseqüentemente, os testes funcionarão, com visto nas simulações da Tabela 16. Ademais, pode-se ver que ao nível de significância de 5% e considerando a decisão da maioria dos testes, houve unanimidade entre os testes para os dados sobre os crânios de gorilas, chimpanzés, orangotangos na categoria fêmea e os dados referente ao dígito 3, ou seja, não rejeita-se a hipótese de que esses dados provêm de uma distribuição Bingham complexa. Por outro lado, houve divergência entre o TKS e os demais testes nos dados de vértebras na categoria pequena, porém pela maioria, também não rejeita-se a hipótese nula. Por fim, ocorreram empates, em que dois testes rejeitam a hipótese nula e os outros dois não rejeitam, essa situação ocorreu nos dados de orangotango macho, vértebra grande e para os dados de conchas. Porém caso fosse considerado um nível de 1%, poderia-se dizer que todos os dados possuem distribuição Bingham complexa.

Tabela 19 – Resultados para os testes de adequação da distribuição Bingham complexa utilizando os testes TRV, TKS, TCM, TAD.

Dados		n	k	λ_1	λ_{k-2}	Valor- p			
						TRV	TKS	TCM	TAD
Gorila	Macho	29	8	-13488	-790	0,4930	0,9810	0,9700	0,9570
	Fêmea	30	8	-16175	-1062	0,9730	0,2000	0,1030	0,1490
Chimpanzé	Macho	28	8	-14136	-955	0,1390	0,2390	0,2620	0,3310
	Fêmea	26	8	-10228	-827	0,1250	0,0410	0,1130	0,0850
Orangotango	Macho	30	8	-6205	-665	0,0430	0,0250	0,0060	0,0050
	Fêmea	24	8	-10017	-1015	0,1730	0,7580	0,4740	0,4960
Vértebras	Grande	23	6	-4862	-446	0,1650	0,0700	0,0110	0,0120
	Pequena	23	6	-5509	-320	0,2670	0,0350	0,0960	0,1480
Dígito 3		30	13	-7955	-22	0,9740	0,3230	0,2610	0,3810
Conchas		21	3	-209	-	0,0280	0,3280	0,2000	0,0390

Na Tabela 20 estão os resultados para os dados do crescimento em 18 ratos com 8 marcos ao longo de 8 períodos. Note que as estimativas dos dois autovalores são muito grandes em valor absoluto, indicando alta concentração em todos os 8 períodos de crescimento.

Considerando a decisão da maioria dos testes, apenas para dos dados referentes ao 90^o dia, que a hipótese de que os dados não provêm da distribuição Bingham complexa é rejeitada. Para os demais dias de crescimento, não rejeita-se a hipótese de que esses dados possuem uma distribuição Bingham complexa ao nível de significância de 5%.

Tabela 20 – Resultados para os testes de adequação da distribuição Bingham complexa para os dados dos crânios de 18 ratos e 8 marcos através de 8 momentos de crescimento utilizando os testes: TRV, TKS, TCM e TAD.

Dias	$\hat{\lambda}_1$	$\hat{\lambda}_{k-2}$	Valor- <i>p</i>			
			TRV	TKS	TCM	TAD
7	-27806	-2290	0,5970	0,5380	0,7310	0,7830
14	-28379	-3272	0,4580	0,3340	0,2290	0,2160
21	-41034	-4925	0,5690	0,1550	0,1010	0,0890
30	-42981	-4520	0,5950	0,8060	0,6990	0,6950
40	-38630	-5273	0,4630	0,7590	0,7320	0,7050
60	-40145	-4348	0,0720	0,1730	0,0430	0,0720
90	-95458	-4198	0,0050	0,1180	0,0210	0,0360
150	-55655	-3488	0,1430	0,0480	0,1420	0,2110

5.8 CONCLUSÃO

Neste Capítulo foi introduzido uma aproximação da distribuição Bingham complexa para a distribuição normal multivariada complexa singular. Essa aproximação foi utilizada em uma forma quadrática que resulta na distribuição Qui-quadrado, que foi utilizada como uma distribuição de referência para avaliar a bondade de ajuste da distribuição Bingham complexa. Essa aproximação foi avaliada para diferentes configurações da matriz de parâmetros, com a matriz de rotação fixa e modificando os autovalores, e viu-se situações em que essa aproximação funcionava. A avaliação dessa aproximação foi feita através dos testes: TRV, TKS, TCM e TAD, no qual concluiu-se que a aproximação é boa para configurações de autovalores que possuem o penúltimo autovalor menor ou igual a -8 .

Essa aproximação foi avaliada para a matriz estimada dos parâmetros via máxima verossimilhança da distribuição Bingham complexa. Entretanto, os resultados não foram satisfatórios, observou-se que os tamanhos dos testes ficavam muito abaixo do valor teórico. Assim, foi utilizada uma transformação nos dados no qual a distribuição conjunta desses dados transformados é Dirichlet simétrica, em que a marginal de cada variável possui distribuição beta. Deste modo, foram utilizados os testes de adequação para a distribuição beta com o método *bootstrap*, em que uma distribuição empírica para cada teste foi gerada e os testes foram realizados normalmente. Ademais, uma análise de poder foi realizada e o desempenho de cada teste aumenta à medida que o tamanho da amostra cresce.

Os testes de adequação foram utilizados em diversos bancos de dados, com o intuito de verificar a bondade de ajuste da distribuição Bingham complexa. Viu-se que a maioria dos dados se ajustam ao modelo e que pode ocorrer divergência entre as decisões dos testes.

6 CONSIDERAÇÕES FINAIS

Várias contribuições foram desenvolvidas para a análise estatística de forma. Especificamente, no Capítulo 3, foram propostos vários classificadores para serem utilizados no espaço das pré-formas com uma variedade de critérios. Ademais, todos os classificadores demonstraram bons desempenhos quando os dados possuem pouca variabilidade (alta concentração), já quando os dados são muito heterogêneos, a SVM demonstrou um desempenho superior aos outros classificadores. Outro ponto importante, foi a combinação entre os classificadores pelo método *ensemble*, o que pode proporcionar uma maior credibilidade na classificação, embora, às vezes a taxa de erro não melhora em comparação com os classificadores individuais.

No Capítulo 4, foi proposto um simulador para a distribuição Watson complexa utilizando uma desigualdade e a distribuição complexa angular central gaussiana. O simulador possui uma alta taxa de aceitação quando os dados não possuem alta concentração e o valor da dimensão do vetor de pré-forma é pequeno. Posteriormente, foi abordado o contexto da bondade de ajuste para a distribuição Watson complexa. Inicialmente, foi utilizada uma aproximação pela distribuição normal multivariada complexa singular, essa aproximação foi utilizada em uma transformação para a distribuição Qui-quadrado, que foi verificada através dos testes da razão de verossimilhanças, Kolmogorov-Smirnov, Cramér-von Mises e Anderson-Darling. Concluiu-se através de simulações de Monte Carlo que a aproximação é boa para $\kappa \geq 10$. Porém, na prática é necessário estimar esse parâmetro e a aproximação não funcionou, pois os tamanhos calculados através da simulação de Monte Carlo ficaram muito abaixo do valor teórico testado. Assim, foi proposta uma nova transformação que resultou em uma distribuição de Dirichlet, no qual produziram bons resultados com a utilização do *bootstrap*. Uma análise de poder também foi realizada para várias configurações dos parâmetros com um esquema de contaminação da amostra. Por fim, esses testes foram utilizados em vários bancos de dados, em que se pôde constatar a adequação da distribuição Watson complexa.

No Capítulo 5, foi utilizada a distribuição Bingham complexa e foram propostos dois geradores. Esses geradores proporcionam um controle maior na simulação, pois o pesquisador poderá fornecer a matriz de parâmetros, ou ainda, apenas os autovalores e a forma média desejada. A taxa de aceitação do simulador depende se os autovalores são pequenos (em valor absoluto) e quando a dimensão do vetor de pré-forma também é baixa. À medida que ambos aumentam, a taxa de aceitação acaba diminuindo. Além disso, foi proposto uma aproximação para a distribuição normal multivariada complexa singular e foi utilizado um resultado que leva

na distribuição Qui-quadrado. Essa aproximação também foi avaliada utilizando os mesmos testes de adequação que foram utilizados para a distribuição Watson complexa. Viu-se que a aproximação é boa para autovalores não muito elevados da matriz de parâmetros. Posteriormente, foi utilizada uma transformação considerando as estimativas dos parâmetros, que forneceu um bom desempenho com relação ao tamanho e poder dos testes. Essa transformação foi utilizada em vários bancos de dados e constatou-se que a grande maioria dos dados utilizados se adequaram a distribuição Bingham complexa.

Comparando a adequação nas bases de dados entre a distribuição Watson e Bingham complexa, observou-se que a maioria dos dados se adequam à distribuição Bingham. Isso ocorreu, devido ao fato de que, a distribuição Bingham complexa possui mais parâmetro e, conseqüentemente, acaba sendo mais flexível. Viu-se também que algumas bases de dados se adequaram em ambas distribuições. Ademais, observou-se que os testes podem divergir quando o tamanho da amostra é pequeno.

Como trabalhos futuros temos, entre outras, as seguintes propostas

- Avaliar os classificadores em transformações no espaço tangente, como: análise de componentes principais e componentes independentes;
- Utilizar os classificadores no espaço dos marcos, na análise de tamanho e forma ou forma;
- Encontrar correções analíticas para os testes de bondade de ajuste quando os parâmetros não são conhecidos.

REFERÊNCIAS

- ALDRICH, J. R. A. Fisher and the making of maximum likelihood 1912-1922. **Statistical science**, Institute of Mathematical Statistics, v. 12, n. 3, p. 162–176, 1997.
- AMARAL, G. A.; DRYDEN, I.; WOOD, A. T. A. Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. **Journal of the American Statistical Association**, Taylor & Francis, v. 102, n. 478, p. 695–707, 2007.
- ANDERSON, C. Object recognition using statistical shape analysis. **PhD, University of Leeds**, 1997.
- ANDERSON, T. W.; DARLING, D. A. A test of goodness of fit. **Journal of the American statistical association**, Taylor & Francis, v. 49, n. 268, p. 765–769, 1954.
- ASSIS, E. C.; SOUZA, R. M. C. R.; AMARAL, J. G. A. Using bagging to enhance clustering procedures for planar shapes. **International Journal of Business Intelligence and Data Mining**, 2018.
- BEST, D.; FISHER, N. Goodness-of-fit and discordancy tests for samples from the Watson distribution on the sphere. **Australian Journal of Statistics**, Wiley Online Library, v. 28, n. 1, p. 13–31, 1986.
- BISHOP, C. M. Pattern recognition and machine learning (information science and statistics) springer-verlag new york. **Inc. Secaucus, NJ, USA**, 2006.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. 2. ed. [S.l.]: SBM, 2010.
- BOOKSTEIN, F. L. A statistical method for biological shape comparisons. **Journal of Theoretical Biology**, Elsevier, v. 107, n. 3, p. 475–520, 1984.
- BOOKSTEIN, F. L. Size and shape spaces for landmark data in two dimensions. **Statistical science**, Institute of Mathematical Statistics, v. 1, n. 2, p. 181–222, 1986.
- BOOKSTEIN, F. L. **Morphometric tools for landmark data: geometry and biology**. [S.l.]: Cambridge University Press, 1997.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. **A training algorithm for optimal margin classifiers**. [S.l.], 1992. 144–152 p.
- BOUBOULIS, P.; THEODORIDOU, E.; THEODORIDIS, S. Complex support vector machines for quaternary classification. In: IEEE. **Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on**. [S.l.], 2013. p. 1–6.
- BOYD, S.; VANDENBERGHE, L. **Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares**. [S.l.]: Cambridge university press, 2018.
- CARVALHO, J. B. de. **Classificador Máquina de Suporte Vetorial com Análise de Fourier Aplicada em Dados de EEG e EMG**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2016.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. [S.l.]: Duxbury Pacific Grove, 2002. v. 2.

CORDEIRO, G. M.; CRIBARI-NETO, F. **An introduction to Bartlett correction and bias reduction**. [S.l.]: Springer, 2014.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

CRAMMER, K.; SINGER, Y. On the algorithmic implementation of multiclass kernel-based vector machines. **Journal of machine learning research**, v. 2, n. Dec, p. 265–292, 2001.

CRISTIANINI, N.; SHAWE-TAYLOR, J. **An introduction to support vector machines and other kernel-based learning methods**. [S.l.]: Cambridge university press, 2000.

CSÖRGŐ, S.; FARAWAY, J. J. The exact and asymptotic distributions of Cramér-von Mises statistics. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 58, n. 1, p. 221–234, 1996.

DEVROYE, L. **Non-uniform random variate generation**. 1. ed. [S.l.]: Springer, 1986.

DRYDEN, I. L. **shapes package**. Vienna, Austria, 2018. Contributed package, Version 1.2.4. Disponível em: <<http://www.R-project.org>>.

DRYDEN, I. L.; MARDIA, K. V. **Statistical Shape Analysis: With Applications in R**. [S.l.]: John Wiley & Sons, 2016.

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

EVERITT, B. S.; LANDAU, S.; LEESE, M.; STAHL, D. **Cluster Analysis, Fifth Edition**. 5th. ed. [S.l.]: Wiley, 2011.

FERGUSON, T. S. A bayesian analysis of some nonparametric problems. **The annals of statistics**, JSTOR, p. 209–230, 1973.

FISCHER, R. F. **Precoding and signal shaping for digital transmission**. [S.l.]: John Wiley & Sons, 2002.

FISHER, N.; BEST, D. Goodness-of-fit tests for Fisher's distribution on the sphere. **Australian Journal of Statistics**, Wiley Online Library, v. 26, n. 2, p. 142–150, 1984.

FISHER, R. A. On an absolute criterion for fitting frequency curves. **Messenger of Mathematics**, v. 41, p. 155–160, 1992.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of computer and system sciences**, Elsevier, v. 55, n. 1, p. 119–139, 1997.

FRIEDMAN, H. P.; RUBIN, J. On some invariant criteria for grouping data. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 62, n. 320, p. 1159–1178, 1967.

GARCÍA, E.; LOZANO, F. Boosting support vector machines. In: **MLDM Posters**. [S.l.: s.n.], 2007. p. 153–167.

GENTLE, J. E. **Numerical linear algebra for applications in statistics**. [S.l.]: Springer, 1998.

GOODALL, C. Procrustes methods in the statistical analysis of shape. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 285–339, 1991.

GOODMAN, N. R. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). **The Annals of mathematical statistics**, JSTOR, v. 34, n. 1, p. 152–177, 1963.

GRATISPNG. Grátispng.com. 2019. Disponível em: <<https://www.gratispng.com/png-5r1suw/>>. Acesso em: 01 set. 2019.

HALD, A. **A history of probability and statistics and their applications before 1750**. [S.l.]: John Wiley & Sons, 2003. v. 501.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. . **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: New York, NY: Springer, 2001.

HORN, C. R. J. R. A. **Matrix Analysis**. [S.l.]: Cambridge University Press, 1990.

HSU, C.-W.; LIN, C.-J. A comparison of methods for multiclass support vector machines. **IEEE transactions on Neural Networks**, IEEE, v. 13, n. 2, p. 415–425, 2002.

HUANG, Z.-Y.; TSAI, P.-Y. Efficient implementation of qr decomposition for gigabit mimo-ofdm systems. **IEEE Transactions on Circuits and Systems I: Regular Papers**, IEEE, v. 58, n. 10, p. 2531–2542, 2011.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. [S.l.]: Springer, 2013. v. 112.

JAYASUMANA, S.; SALZMANN, M.; LI, H.; HARANDI, M. A framework for shape analysis via hilbert space embedding. In: **Proceedings of the IEEE International Conference on Computer Vision**. [S.l.: s.n.], 2013. p. 1249–1256.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis (6th Edition)**. 6. ed. [S.l.]: Prentice Hall, 2007.

KENDALL, D. G. The diffusion of shape. **Advances in applied probability**, Cambridge University Press, v. 9, n. 3, p. 428–430, 1977.

KENDALL, D. G. Shape manifolds, procrustean metrics, and complex projective spaces. **Bulletin of the London Mathematical Society**, Wiley Online Library, v. 16, n. 2, p. 81–121, 1984.

KENDALL, D. G.; BARDEN, D.; CARNE, T. K.; LE, H. **Shape and shape theory**. [S.l.]: John Wiley & Sons, 2009. v. 500.

KENT, J. T. The complex Bingham distribution and shape analysis. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 285–299, 1994.

KENT, J. T.; CONSTABLE, P. D.; ER, F. Simulation for the complex Bingham distribution. **Statistics and Computing**, Springer, v. 14, n. 1, p. 53–57, 2004.

KENT, J. T.; GANEIBER, A. M.; MARDIA, K. V. A new method to simulate the Bingham and related distributions in directional data analysis with applications. **arXiv preprint arXiv:1310.8110**, 2013.

KHATRI, C. G. Some results for the singular normal multivariate regression models. **Sankhyā: The Indian Journal of Statistics, Series A**, JSTOR, p. 267–280, 1968.

- KOLMOGOROV, A. Sulla determinazione empirica di una legge di distribuzione. **Inst. Ital. Attuari, Giorn.**, v. 4, p. 83–91, 1933.
- KUNCHEVA, L. I. **Combining pattern classifiers: methods and algorithms**. [S.l.]: John Wiley & Sons, 2004.
- LEY, C.; VERDEBOUT, T. **Modern directional statistics**. [S.l.]: Chapman and Hall/CRC, 2017.
- LOHMANN, G. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. **Journal of the International Association for Mathematical Geology**, Springer, v. 15, n. 6, p. 659–672, 1983.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.
- MARDIA, K.; DRYDEN, I. The complex Watson distribution and shape analysis. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 61, n. 4, p. 913–926, 1999.
- MARDIA, K.; KENT, J.; BIBBY, J. **Multivariate analysis**. [S.l.]: Academic Press, London, 1979.
- MARDIA, K. V.; JUPP, P. E. **Directional statistics**. [S.l.]: John Wiley & Sons, 2000.
- MIRSKY, L. **An Introduction to Linear Algebra**. First. [S.l.]: Oxford U. Press, 1955.
- OLIVEIRA, R. A. de. **Algoritmos para Determinação do Número de Grupos em Estudos de Formas Planas**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2016. <<https://www3.ufpe.br/ppge/images/dissertacoes/dissert148.pdf>>.
- PLATT, J. C.; CRISTIANINI, N.; SHAWE-TAYLOR, J. Large margin dags for multiclass classification. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2000. p. 547–553.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>.
- RAO, C. R. **Linear statistical inference and its applications**. [S.l.]: Wiley New York, 1973. v. 2.
- RENCHER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. [S.l.]: John Wiley & Sons, 2008.
- RUBIN, J. Optimal classification into groups: an approach for solving the taxonomy problem. **Journal of Theoretical Biology**, Elsevier, v. 15, n. 1, p. 103–144, 1967.
- SANTOS, E. M. dos. **Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência**. Tese (Doutorado) — Universidade Federal da Paraíba, 2002.
- SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. **Tübingen: Max-Planck-Institut für biologische Kybernetik**, 1996.
- SEBER, G. A. F. **Multivariate observations**. [S.l.]: John Wiley & Sons, 1984. v. 252.

- SILVERMAN, B. W. **Density estimation for statistics and data analysis**. [S.l.]: CRC press, 1986. v. 26.
- SMALL, C. G. **The statistical theory of shape**. [S.l.]: Springer-Verlag, New York, 1996.
- SMIRNOV, N. V. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. **Bull. Math. Univ. Moscou**, v. 2, n. 2, p. 3–14, 1939.
- SOUTHWORTH, R.; MARDIA, K.; TAYLOR, C. Transformation-and label-invariant neural network for the classification of landmark data. **Journal of Applied Statistics**, Taylor & Francis, v. 27, n. 2, p. 205–215, 2000.
- SOUZA, M. P. S.; FILHO, T. M. S.; AMARAL, G. J. A.; SOUZA, R. M. C. R. Investigating different fitness criteria for swarm-based clustering. **International Journal of Business Intelligence and Data Mining**, 2017.
- STEPHENS, M. A. Edf statistics for goodness of fit and some comparisons. **Journal of the American statistical Association**, Taylor & Francis Group, v. 69, n. 347, p. 730–737, 1974.
- TSUKUMA, H.; KUBOKAWA, T. Estimation of the mean vector in a singular multivariate normal distribution. **Journal of Multivariate Analysis**, Elsevier, v. 140, p. 245–258, 2015.
- TYLER, D. E. Statistical analysis for the angular central gaussian distribution on the sphere. **Biometrika**, Oxford University Press, v. 74, n. 3, p. 579–589, 1987.
- WESTON, J.; WATKINS, C. **Multi-class support vector machines**. [S.l.], 1998.
- WOODING, R. A. The multivariate distribution of complex normal variables. **Biometrika**, JSTOR, v. 43, n. 1/2, p. 212–215, 1956.
- ZHOU, Z.-H. **Ensemble methods: foundations and algorithms**. [S.l.]: Chapman and Hall/-CRC, 2012.

APÊNDICE A – SIMULAÇÕES DOS TESTES DE BONDADE DE AJUSTE - WATSON COMPLEXA

NÍVEL NOMINAL - PARÂMETROS CONHECIDOS - WATSON COMPLEXA

Nas Tabelas 21 e 22 estão as simulações para os tamanhos de 1% e 5% quando os parâmetros são conhecidos, respectivamente.

Tabela 21 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\mu = (0, 0, 0, 0, 1)^T$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.

Teste	n	κ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	0,9998	0,1426	0,0094	0,0076	0,0092	0,0106	0,0066	0,0090	0,0102	0,0116	0,0084
	30	1,0000	1,0000	0,3116	0,0142	0,0088	0,0108	0,0122	0,0102	0,0096	0,0106	0,0094	0,0094
	40	1,0000	1,0000	0,4724	0,0132	0,0098	0,0096	0,0106	0,0112	0,0094	0,0092	0,0092	0,0096
	50	1,0000	1,0000	0,6386	0,0156	0,0086	0,0112	0,0104	0,0076	0,0092	0,0100	0,0082	0,0124
	100	1,0000	1,0000	0,9782	0,0262	0,0066	0,0122	0,0082	0,0090	0,0098	0,0096	0,0108	0,0080
	200	1,0000	1,0000	1,0000	0,0572	0,0116	0,0102	0,0104	0,0124	0,0114	0,0086	0,0098	0,0084
TKS	20	1,0000	1,0000	0,2106	0,0126	0,0118	0,0106	0,0112	0,0106	0,0076	0,0122	0,0096	0,0086
	30	1,0000	1,0000	0,5870	0,0162	0,0096	0,0108	0,0126	0,0090	0,0102	0,0108	0,0072	0,0072
	40	1,0000	1,0000	1,0000	0,0132	0,0090	0,0084	0,0104	0,0108	0,0100	0,0090	0,0094	0,0104
	50	1,0000	1,0000	1,0000	0,0102	0,0102	0,0088	0,0106	0,0100	0,0102	0,0094	0,0098	0,0102
	100	1,0000	1,0000	1,0000	0,0154	0,0074	0,0098	0,0072	0,0080	0,0090	0,0120	0,0092	0,0082
	200	1,0000	1,0000	1,0000	0,0324	0,0108	0,0078	0,0076	0,0120	0,0102	0,0100	0,0080	0,0070
TCM	20	1,0000	1,0000	0,2594	0,0120	0,0106	0,0112	0,0108	0,0104	0,0072	0,0132	0,0102	0,0098
	30	1,0000	1,0000	0,5416	0,0132	0,0094	0,0096	0,0120	0,0104	0,0092	0,0106	0,0076	0,0102
	40	1,0000	1,0000	0,7918	0,0148	0,0084	0,0076	0,0106	0,0124	0,0094	0,0088	0,0080	0,0102
	50	1,0000	1,0000	0,9508	0,0142	0,0112	0,0094	0,0112	0,0080	0,0096	0,0084	0,0072	0,0106
	100	1,0000	1,0000	1,0000	0,0240	0,0078	0,0112	0,0084	0,0094	0,0122	0,0096	0,0110	0,0086
	200	1,0000	1,0000	1,0000	0,0470	0,0124	0,0072	0,0102	0,0118	0,0116	0,0104	0,0112	0,0090
TAD	20	1,0000	1,0000	0,2256	0,0112	0,0102	0,0102	0,0104	0,0092	0,0086	0,0114	0,0112	0,0100
	30	1,0000	1,0000	0,5352	0,0132	0,0102	0,0088	0,0118	0,0096	0,0088	0,0100	0,0082	0,0112
	40	1,0000	1,0000	0,8170	0,0138	0,0084	0,0080	0,0112	0,0116	0,0086	0,0078	0,0074	0,0104
	50	1,0000	1,0000	0,9772	0,0134	0,0114	0,0102	0,0102	0,0068	0,0086	0,0086	0,0076	0,0122
	100	1,0000	1,0000	1,0000	0,0282	0,0074	0,0114	0,0090	0,0100	0,0108	0,0098	0,0100	0,0084
	200	1,0000	1,0000	1,0000	0,0696	0,0128	0,0078	0,0092	0,0124	0,0118	0,0086	0,0114	0,0086

Tabela 22 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD, considerando a forma média $\mu = (0, 0, 0, 0, 1)^T$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.

Teste	n	κ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	1,0000	0,5674	0,0948	0,0898	0,0982	0,1030	0,0936	0,1046	0,0978	0,1046	0,0894
	30	1,0000	1,0000	0,7726	0,1078	0,0958	0,1022	0,1008	0,0970	0,1076	0,1012	0,0866	0,0942
	40	1,0000	1,0000	0,8838	0,1214	0,0990	0,0950	0,1020	0,0990	0,1002	0,1026	0,1018	0,0938
	50	1,0000	1,0000	0,9536	0,1280	0,1038	0,1050	0,1012	0,1016	0,0992	0,0956	0,0992	0,1022
	100	1,0000	1,0000	0,9998	0,1770	0,0906	0,1046	0,0908	0,0986	0,1026	0,1002	0,0962	0,0986
	200	1,0000	1,0000	1,0000	0,2932	0,1056	0,0956	0,1066	0,1002	0,1068	0,1036	0,1026	0,0954
TKS	20	1,0000	1,0000	1,0000	0,1036	0,0974	0,0938	0,0966	0,0966	0,0988	0,1026	0,1012	0,0896
	30	1,0000	1,0000	1,0000	0,1200	0,0930	0,1010	0,1006	0,1066	0,1024	0,0990	0,0928	0,0996
	40	1,0000	1,0000	1,0000	0,1146	0,0946	0,0932	0,1012	0,1026	0,0942	0,0978	0,1012	0,0956
	50	1,0000	1,0000	1,0000	0,1158	0,1040	0,0998	0,1034	0,1036	0,0948	0,0952	0,0942	0,0998
	100	1,0000	1,0000	1,0000	0,1276	0,0864	0,0932	0,0872	0,0936	0,0936	0,0892	0,0872	0,0860
	200	1,0000	1,0000	1,0000	0,2132	0,0982	0,0888	0,0994	0,0980	0,0980	0,0914	0,0960	0,0880
TCM	20	1,0000	1,0000	0,8198	0,1064	0,0950	0,0920	0,1034	0,0954	0,0988	0,1044	0,1004	0,0912
	30	1,0000	1,0000	0,9824	0,1190	0,0946	0,1024	0,1010	0,1020	0,1064	0,1024	0,0932	0,0920
	40	1,0000	1,0000	0,9998	0,1274	0,1016	0,0926	0,1028	0,0974	0,0958	0,0980	0,0964	0,0894
	50	1,0000	1,0000	1,0000	0,1254	0,1048	0,1014	0,1052	0,1080	0,0952	0,0996	0,0966	0,1010
	100	1,0000	1,0000	1,0000	0,1558	0,0896	0,1012	0,0966	0,1052	0,0994	0,0958	0,0982	0,0910
	200	1,0000	1,0000	1,0000	0,2612	0,1010	0,0928	0,1028	0,1022	0,1046	0,0948	0,1012	0,0938
TAD	20	1,0000	1,0000	0,8252	0,0992	0,0886	0,0940	0,1024	0,0962	0,0972	0,1034	0,0998	0,0906
	30	1,0000	1,0000	0,9952	0,1154	0,0888	0,1042	0,1020	0,1036	0,1088	0,1020	0,0900	0,0974
	40	1,0000	1,0000	1,0000	0,1306	0,0964	0,0944	0,1014	0,0996	0,0988	0,0984	0,0992	0,0928
	50	1,0000	1,0000	1,0000	0,1302	0,0970	0,0988	0,1036	0,1056	0,0942	0,0974	0,0950	0,1024
	100	1,0000	1,0000	1,0000	0,1870	0,0912	0,1012	0,0988	0,1010	0,0998	0,0960	0,0976	0,0932
	200	1,0000	1,0000	1,0000	0,3742	0,1062	0,0968	0,1064	0,1022	0,1066	0,0970	0,1040	0,0912

NÍVEL NOMINAL - PARÂMETROS ESTIMADOS - WATSON COMPLEXA

Nas Tabelas 23, 24 e 25 estão as simulações para os tamanhos de 1%, 5% e 10% quando os parâmetros da distribuição Watson complexa são desconhecidos, respectivamente.

Tabela 23 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, considerando a forma média $\mu = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.

Teste	n	κ							
		10	20	30	40	50	100	1000	5000
TRV	20	0,0088	0,0102	0,0124	0,0102	0,0106	0,0106	0,0116	0,0124
	30	0,0096	0,0092	0,0128	0,0100	0,0126	0,0108	0,0098	0,0108
	40	0,0066	0,0104	0,0080	0,0098	0,0114	0,0128	0,0086	0,0090
	50	0,0096	0,0080	0,0090	0,0116	0,0096	0,0100	0,0066	0,0108
	100	0,0070	0,0090	0,0112	0,0108	0,0084	0,0092	0,0078	0,0104
	200	0,0160	0,0088	0,0106	0,0090	0,0100	0,0104	0,0122	0,0068
TKS	20	0,0116	0,0122	0,0104	0,0090	0,0106	0,0090	0,0092	0,0110
	30	0,0072	0,0108	0,0114	0,0108	0,0122	0,0098	0,0096	0,0100
	40	0,0082	0,0096	0,0090	0,0130	0,0118	0,0124	0,0110	0,0114
	50	0,0078	0,0112	0,0114	0,0098	0,0112	0,0096	0,0098	0,0124
	100	0,0082	0,0108	0,0090	0,0102	0,0106	0,0084	0,0098	0,0106
	200	0,0134	0,0114	0,0098	0,0102	0,0086	0,0132	0,0108	0,0090
TCM	20	0,0110	0,0108	0,0114	0,0110	0,0094	0,0078	0,0086	0,0098
	30	0,0092	0,0104	0,0106	0,0108	0,0128	0,0118	0,0086	0,0088
	40	0,0084	0,0094	0,0120	0,0122	0,0114	0,0094	0,0094	0,0100
	50	0,0096	0,0090	0,0116	0,0104	0,0094	0,0108	0,0096	0,0112
	100	0,0066	0,0120	0,0080	0,0104	0,0110	0,0086	0,0096	0,0104
	200	0,0126	0,0102	0,0086	0,0096	0,0094	0,0108	0,0090	0,0076
TAD	20	0,0120	0,0112	0,0130	0,0100	0,0108	0,0084	0,0090	0,0116
	30	0,0100	0,0108	0,0132	0,0116	0,0114	0,0112	0,0098	0,0090
	40	0,0072	0,0124	0,0128	0,0102	0,0126	0,0090	0,0088	0,0100
	50	0,0104	0,0098	0,0124	0,0126	0,0106	0,0102	0,0086	0,0120
	100	0,0054	0,0102	0,0108	0,0092	0,0102	0,0084	0,0096	0,0104
	200	0,0118	0,0106	0,0094	0,0116	0,0092	0,0118	0,0106	0,0068

Tabela 24 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, considerando a forma média $\mu = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 5%.

Teste	n	κ							
		10	20	30	40	50	100	1000	5000
TRV	20	0,0424	0,0478	0,0572	0,0504	0,0536	0,0502	0,0568	0,0506
	30	0,0406	0,0458	0,0570	0,0462	0,0514	0,0482	0,0514	0,0484
	40	0,0402	0,0564	0,0524	0,0566	0,0504	0,0516	0,0538	0,0566
	50	0,0466	0,0504	0,0482	0,0550	0,0494	0,0538	0,0488	0,0518
	100	0,0514	0,0520	0,0532	0,0524	0,0448	0,0462	0,0460	0,0496
	200	0,0710	0,0458	0,0512	0,0514	0,0496	0,0476	0,0482	0,0442
TKS	20	0,0444	0,0518	0,0530	0,0510	0,0488	0,0508	0,0538	0,0484
	30	0,0424	0,0504	0,0554	0,0558	0,0528	0,0450	0,0554	0,0470
	40	0,0476	0,0522	0,0528	0,0550	0,0518	0,0502	0,0462	0,0540
	50	0,0462	0,0476	0,0588	0,0510	0,0484	0,0502	0,0502	0,0538
	100	0,0428	0,0498	0,0442	0,0488	0,0540	0,0462	0,0510	0,0540
	200	0,0638	0,0462	0,0512	0,0510	0,0442	0,0550	0,0492	0,0516
TCM	20	0,0466	0,0550	0,0550	0,0504	0,0534	0,0494	0,0516	0,0524
	30	0,0418	0,0504	0,0546	0,0532	0,0566	0,0480	0,0526	0,0504
	40	0,0454	0,0534	0,0522	0,0530	0,0538	0,0536	0,0452	0,0530
	50	0,0454	0,0484	0,0550	0,0480	0,0506	0,0544	0,0486	0,0528
	100	0,0434	0,0504	0,0490	0,0482	0,0586	0,0450	0,0548	0,0500
	200	0,0620	0,0478	0,0490	0,0518	0,0476	0,0532	0,0478	0,0492
TAD	20	0,0492	0,0588	0,0582	0,0504	0,0506	0,0488	0,0536	0,0518
	30	0,0428	0,0522	0,0556	0,0498	0,0544	0,0468	0,0492	0,0484
	40	0,0408	0,0562	0,0536	0,0552	0,0510	0,0540	0,0498	0,0534
	50	0,0456	0,0568	0,0540	0,0562	0,0478	0,0532	0,0514	0,0544
	100	0,0408	0,0494	0,0496	0,0464	0,0564	0,0472	0,0488	0,0516
	200	0,0566	0,0488	0,0506	0,0506	0,0498	0,0518	0,0504	0,0478

Tabela 25 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, considerando a forma média $\mu = (0, 0, 0, 0, 1)^\top$, diferentes valores de κ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.

Teste	n	κ							
		10	20	30	40	50	100	1000	5000
TRV	20	0,0942	0,0960	0,1076	0,1050	0,1024	0,0984	0,1076	0,1010
	30	0,0892	0,0998	0,1084	0,0972	0,1012	0,1008	0,1046	0,0986
	40	0,0874	0,1086	0,1042	0,1034	0,0948	0,1052	0,1074	0,1062
	50	0,0938	0,1002	0,0980	0,1096	0,0992	0,0986	0,0990	0,1018
	100	0,1054	0,0966	0,1032	0,1024	0,0924	0,0972	0,1022	0,1082
	200	0,1268	0,0918	0,1028	0,1008	0,0958	0,0968	0,0956	0,0956
TKS	20	0,0902	0,1000	0,1006	0,1006	0,0998	0,1016	0,1008	0,1022
	30	0,0864	0,0982	0,1030	0,1046	0,1060	0,0998	0,1076	0,0962
	40	0,0924	0,1030	0,1054	0,0990	0,1002	0,1048	0,0996	0,1040
	50	0,0934	0,0950	0,1120	0,1000	0,0988	0,0996	0,1062	0,1040
	100	0,0918	0,0970	0,0948	0,1016	0,1038	0,0980	0,1000	0,1034
	200	0,1156	0,0990	0,0978	0,1014	0,0986	0,1010	0,0968	0,1048
TCM	20	0,0934	0,1066	0,1022	0,1022	0,1028	0,0986	0,0988	0,1022
	30	0,0884	0,1052	0,1056	0,1032	0,1024	0,0950	0,1034	0,1032
	40	0,0908	0,1076	0,1046	0,1022	0,1022	0,1088	0,0948	0,1084
	50	0,0916	0,1020	0,1084	0,0990	0,0942	0,1052	0,1052	0,1056
	100	0,0910	0,0986	0,0990	0,0954	0,1072	0,0992	0,0978	0,1086
	200	0,1150	0,0994	0,0962	0,1016	0,1010	0,1000	0,0980	0,0990
TAD	20	0,1000	0,1104	0,1048	0,0996	0,0990	0,1010	0,1040	0,1044
	30	0,0880	0,1070	0,1086	0,1056	0,1048	0,0954	0,0982	0,1004
	40	0,0878	0,1108	0,1106	0,1112	0,1036	0,1076	0,0948	0,1098
	50	0,0948	0,1088	0,1052	0,1058	0,0948	0,1056	0,1048	0,1082
	100	0,0892	0,1022	0,1036	0,1000	0,1074	0,0946	0,1016	0,1034
	200	0,1110	0,0962	0,0968	0,1008	0,1010	0,0990	0,1022	0,0988

ANÁLISE DO PODER - WATSON COMPLEXA

Nível nominal de $\alpha = 1\%$ **Tabela 26 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.**

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0062	0,0078	0,0060	0,0074	0,0066	0,0084	0,0078
	30	0,0090	0,0098	0,0082	0,0088	0,0064	0,0066	0,0082
	40	0,0084	0,0110	0,0072	0,0074	0,0074	0,0096	0,0100
	50	0,0058	0,0092	0,0098	0,0102	0,0092	0,0110	0,0100
	100	0,0084	0,0114	0,0124	0,0092	0,0122	0,0124	0,0146
	200	0,0178	0,0160	0,0176	0,0178	0,0212	0,0212	0,0222
TKS	20	0,0072	0,0080	0,0084	0,0074	0,0078	0,0088	0,0084
	30	0,0096	0,0094	0,0090	0,0084	0,0052	0,0072	0,0094
	40	0,0076	0,0086	0,0094	0,0068	0,0092	0,0080	0,0090
	50	0,0078	0,0076	0,0084	0,0084	0,0088	0,0090	0,0078
	100	0,0066	0,0120	0,0094	0,0102	0,0096	0,0110	0,0102
	200	0,0132	0,0130	0,0124	0,0142	0,0156	0,0156	0,0156
TCM	20	0,0096	0,0060	0,0080	0,0086	0,0092	0,0088	0,0100
	30	0,0108	0,0096	0,0094	0,0094	0,0062	0,0094	0,0090
	40	0,0078	0,0068	0,0094	0,0070	0,0080	0,0088	0,0090
	50	0,0076	0,0080	0,0108	0,0092	0,0080	0,0116	0,0082
	100	0,0058	0,0110	0,0074	0,0080	0,0098	0,0096	0,0094
	200	0,0116	0,0112	0,0152	0,0118	0,0174	0,0174	0,0148
TAD	20	0,0084	0,0086	0,0094	0,0068	0,0100	0,0084	0,0098
	30	0,0100	0,0088	0,0084	0,0096	0,0070	0,0082	0,0102
	40	0,0078	0,0078	0,0080	0,0068	0,0076	0,0084	0,0090
	50	0,0068	0,0072	0,0128	0,0092	0,0078	0,0104	0,0074
	100	0,0070	0,0106	0,0088	0,0076	0,0072	0,0086	0,0090
	200	0,0112	0,0108	0,0126	0,0098	0,0154	0,0158	0,0138

Tabela 27 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0110	0,0066	0,0098	0,0104	0,0090	0,0106	0,0114
	30	0,0100	0,0100	0,0092	0,0112	0,0120	0,0080	0,0108
	40	0,0100	0,0080	0,0090	0,0104	0,0082	0,0096	0,0114
	50	0,0084	0,0092	0,0100	0,0102	0,0082	0,0104	0,0134
	100	0,0088	0,0118	0,0080	0,0106	0,0076	0,0102	0,0112
	200	0,0104	0,0118	0,0120	0,0086	0,0098	0,0124	0,0132
TKS	20	0,0110	0,0098	0,0118	0,0112	0,0134	0,0094	0,0144
	30	0,0130	0,0114	0,0086	0,0106	0,0116	0,0114	0,0124
	40	0,0088	0,0104	0,0094	0,0118	0,0108	0,0112	0,0120
	50	0,0088	0,0116	0,0112	0,0090	0,0132	0,0122	0,0122
	100	0,0106	0,0118	0,0106	0,0088	0,0092	0,0120	0,0140
	200	0,0122	0,0088	0,0100	0,0108	0,0098	0,0100	0,0120
TCM	20	0,0110	0,0102	0,0124	0,0110	0,0136	0,0128	0,0130
	30	0,0116	0,0134	0,0066	0,0100	0,0118	0,0132	0,0138
	40	0,0072	0,0108	0,0104	0,0116	0,0108	0,0108	0,0134
	50	0,0082	0,0112	0,0116	0,0114	0,0106	0,0132	0,0126
	100	0,0096	0,0092	0,0120	0,0084	0,0108	0,0120	0,0136
	200	0,0108	0,0094	0,0106	0,0118	0,0090	0,0104	0,0128
TAD	20	0,0120	0,0124	0,0118	0,0120	0,0124	0,0144	0,0178
	30	0,0116	0,0114	0,0068	0,0124	0,0146	0,0138	0,0188
	40	0,0088	0,0084	0,0094	0,0126	0,0124	0,0144	0,0148
	50	0,0092	0,0090	0,0132	0,0136	0,0094	0,0144	0,0152
	100	0,0094	0,0084	0,0110	0,0100	0,0114	0,0124	0,0152
	200	0,0116	0,0084	0,0108	0,0110	0,0094	0,0124	0,0140

Tabela 28 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0088	0,0132	0,0094	0,0106	0,0078	0,0100	0,0122
	30	0,0102	0,0072	0,0110	0,0124	0,0106	0,0124	0,0138
	40	0,0120	0,0122	0,0088	0,0112	0,0108	0,0130	0,0158
	50	0,0096	0,0092	0,0108	0,0080	0,0116	0,0102	0,0202
	100	0,0112	0,0132	0,0084	0,0102	0,0120	0,0148	0,0236
	200	0,0096	0,0100	0,0094	0,0080	0,0112	0,0196	0,0372
TKS	20	0,0104	0,0116	0,0086	0,0124	0,0122	0,0154	0,0194
	30	0,0088	0,0094	0,0108	0,0130	0,0140	0,0150	0,0148
	40	0,0108	0,0086	0,0102	0,0106	0,0116	0,0170	0,0178
	50	0,0102	0,0118	0,0102	0,0120	0,0132	0,0146	0,0206
	100	0,0108	0,0080	0,0128	0,0098	0,0150	0,0130	0,0212
	200	0,0092	0,0088	0,0090	0,0098	0,0108	0,0202	0,0334
TCM	20	0,0112	0,0146	0,0078	0,0118	0,0120	0,0182	0,0198
	30	0,0108	0,0098	0,0120	0,0108	0,0138	0,0152	0,0170
	40	0,0100	0,0118	0,0102	0,0108	0,0110	0,0168	0,0176
	50	0,0088	0,0108	0,0112	0,0096	0,0138	0,0118	0,0218
	100	0,0102	0,0102	0,0096	0,0110	0,0154	0,0182	0,0258
	200	0,0078	0,0090	0,0076	0,0102	0,0116	0,0216	0,0428
TAD	20	0,0140	0,0130	0,0086	0,0136	0,0130	0,0170	0,0248
	30	0,0096	0,0090	0,0118	0,0132	0,0154	0,0184	0,0226
	40	0,0112	0,0110	0,0112	0,0128	0,0146	0,0208	0,0260
	50	0,0076	0,0118	0,0084	0,0112	0,0144	0,0150	0,0272
	100	0,0126	0,0096	0,0094	0,0124	0,0174	0,0200	0,0320
	200	0,0088	0,0090	0,0086	0,0114	0,0140	0,0242	0,0468

Tabela 29 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0096	0,0094	0,0076	0,0088	0,0118	0,0144	0,0242
	30	0,0092	0,0098	0,0128	0,0072	0,0130	0,0194	0,0246
	40	0,0108	0,0108	0,0110	0,0102	0,0114	0,0180	0,0336
	50	0,0104	0,0108	0,0102	0,0088	0,0108	0,0222	0,0370
	100	0,0084	0,0076	0,0116	0,0118	0,0150	0,0324	0,0680
	200	0,0114	0,0092	0,0098	0,0120	0,0246	0,0610	0,1406
TKS	20	0,0084	0,0078	0,0120	0,0114	0,0136	0,0198	0,0252
	30	0,0132	0,0080	0,0152	0,0106	0,0142	0,0246	0,0308
	40	0,0100	0,0132	0,0094	0,0126	0,0136	0,0218	0,0378
	50	0,0108	0,0118	0,0132	0,0118	0,0156	0,0238	0,0388
	100	0,0098	0,0094	0,0114	0,0090	0,0170	0,0366	0,0576
	200	0,0108	0,0090	0,0096	0,0118	0,0220	0,0492	0,1056
TCM	20	0,0084	0,0072	0,0120	0,0116	0,0160	0,0212	0,0294
	30	0,0142	0,0090	0,0116	0,0100	0,0160	0,0268	0,0358
	40	0,0096	0,0106	0,0098	0,0130	0,0150	0,0246	0,0450
	50	0,0104	0,0114	0,0128	0,0110	0,0148	0,0258	0,0444
	100	0,0094	0,0090	0,0106	0,0106	0,0186	0,0372	0,0714
	200	0,0102	0,0112	0,0106	0,0118	0,0240	0,0614	0,1476
TAD	20	0,0086	0,0074	0,0124	0,0142	0,0202	0,0262	0,0412
	30	0,0136	0,0074	0,0134	0,0120	0,0210	0,0322	0,0476
	40	0,0104	0,0118	0,0090	0,0124	0,0188	0,0324	0,0540
	50	0,0112	0,0092	0,0146	0,0132	0,0154	0,0338	0,0544
	100	0,0086	0,0094	0,0106	0,0138	0,0208	0,0440	0,0870
	200	0,0122	0,0104	0,0100	0,0126	0,0286	0,0694	0,1696

Tabela 30 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0112	0,0076	0,0114	0,0108	0,0144	0,0218	0,0380
	30	0,0100	0,0064	0,0098	0,0114	0,0214	0,0254	0,0474
	40	0,0100	0,0124	0,0116	0,0132	0,0172	0,0372	0,0680
	50	0,0098	0,0082	0,0092	0,0126	0,0246	0,0398	0,0840
	100	0,0086	0,0100	0,0106	0,0154	0,0332	0,0746	0,1640
	200	0,0108	0,0086	0,0112	0,0204	0,0568	0,1588	0,3756
TKS	20	0,0082	0,0108	0,0114	0,0128	0,0192	0,0316	0,0516
	30	0,0092	0,0084	0,0098	0,0170	0,0214	0,0274	0,0576
	40	0,0114	0,0122	0,0134	0,0114	0,0232	0,0346	0,0712
	50	0,0098	0,0094	0,0100	0,0164	0,0258	0,0438	0,0796
	100	0,0098	0,0080	0,0096	0,0158	0,0304	0,0702	0,1406
	200	0,0080	0,0108	0,0110	0,0216	0,0478	0,1250	0,2940
TCM	20	0,0082	0,0120	0,0124	0,0138	0,0182	0,0350	0,0612
	30	0,0114	0,0080	0,0114	0,0148	0,0266	0,0348	0,0682
	40	0,0100	0,0128	0,0124	0,0138	0,0230	0,0416	0,0836
	50	0,0102	0,0076	0,0106	0,0162	0,0250	0,0536	0,1034
	100	0,0086	0,0104	0,0096	0,0148	0,0342	0,0856	0,1976
	200	0,0104	0,0104	0,0124	0,0238	0,0596	0,1692	0,4036
TAD	20	0,0080	0,0106	0,0140	0,0162	0,0250	0,0468	0,0710
	30	0,0114	0,0082	0,0118	0,0208	0,0340	0,0460	0,0846
	40	0,0116	0,0130	0,0120	0,0158	0,0286	0,0548	0,1070
	50	0,0114	0,0102	0,0120	0,0194	0,0348	0,0650	0,1260
	100	0,0074	0,0114	0,0136	0,0186	0,0396	0,1020	0,2302
	200	0,0094	0,0106	0,0122	0,0266	0,0704	0,1978	0,4628

Tabela 31 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0112	0,0094	0,0112	0,0232	0,0722	0,1504	0,2742
	30	0,0092	0,0090	0,0144	0,0332	0,1024	0,2338	0,4382
	40	0,0100	0,0100	0,0142	0,0496	0,1356	0,3380	0,5688
	50	0,0092	0,0098	0,0150	0,0618	0,1842	0,4144	0,6970
	100	0,0112	0,0130	0,0236	0,1136	0,3842	0,7662	0,9538
	200	0,0108	0,0144	0,0434	0,2472	0,7444	0,9796	0,9996
TKS	20	0,0106	0,0166	0,0180	0,0334	0,0942	0,1822	0,3106
	30	0,0096	0,0134	0,0162	0,0434	0,1144	0,2424	0,4416
	40	0,0096	0,0098	0,0190	0,0486	0,1380	0,3210	0,5496
	50	0,0100	0,0084	0,0184	0,0594	0,1626	0,3814	0,6402
	100	0,0112	0,0128	0,0274	0,0982	0,3278	0,6866	0,9262
	200	0,0104	0,0132	0,0356	0,1904	0,6428	0,9526	0,9988
TCM	20	0,0122	0,0132	0,0192	0,0342	0,1182	0,2300	0,3972
	30	0,0108	0,0110	0,0206	0,0468	0,1458	0,3126	0,5670
	40	0,0094	0,0104	0,0198	0,0570	0,1784	0,4238	0,6966
	50	0,0100	0,0100	0,0182	0,0740	0,2236	0,5096	0,7886
	100	0,0106	0,0142	0,0290	0,1276	0,4386	0,8336	0,9820
	200	0,0114	0,0158	0,0454	0,2722	0,7936	0,9934	0,9998
TAD	20	0,0128	0,0130	0,0216	0,0420	0,1452	0,2686	0,4580
	30	0,0108	0,0122	0,0248	0,0594	0,1740	0,3640	0,6262
	40	0,0102	0,0132	0,0250	0,0686	0,2124	0,4850	0,7498
	50	0,0096	0,0094	0,0234	0,0862	0,2704	0,5616	0,8348
	100	0,0116	0,0142	0,0352	0,1570	0,5002	0,8682	0,9874
	200	0,0100	0,0176	0,0518	0,3074	0,8392	0,9970	1,0000

Nível nominal de $\alpha = 5\%$

Tabela 34 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0418	0,0390	0,0416	0,0428	0,0432	0,0460	0,0446
	30	0,0484	0,0454	0,0430	0,0422	0,0404	0,0404	0,0502
	40	0,0472	0,0468	0,0436	0,0438	0,0436	0,0506	0,0506
	50	0,0466	0,0466	0,0490	0,0468	0,0482	0,0490	0,0476
	100	0,0520	0,0520	0,0544	0,0500	0,0574	0,0612	0,0598
	200	0,0768	0,0750	0,0720	0,0776	0,0844	0,0852	0,0852
TKS	20	0,0444	0,0412	0,0430	0,0406	0,0472	0,0428	0,0432
	30	0,0450	0,0404	0,0458	0,0420	0,0444	0,0438	0,0442
	40	0,0446	0,0486	0,0436	0,0414	0,0454	0,0454	0,0478
	50	0,0422	0,0402	0,0454	0,0454	0,0464	0,0436	0,0492
	100	0,0468	0,0548	0,0524	0,0490	0,0498	0,0512	0,0528
	200	0,0600	0,0588	0,0628	0,0614	0,0686	0,0676	0,0692
TCM	20	0,0456	0,0416	0,0418	0,0404	0,0440	0,0436	0,0452
	30	0,0460	0,0430	0,0434	0,0434	0,0432	0,0440	0,0462
	40	0,0498	0,0462	0,0396	0,0376	0,0434	0,0442	0,0486
	50	0,0466	0,0418	0,0468	0,0436	0,0410	0,0460	0,0498
	100	0,0450	0,0540	0,0504	0,0504	0,0450	0,0494	0,0550
	200	0,0598	0,0612	0,0636	0,0574	0,0704	0,0726	0,0702
TAD	20	0,0474	0,0462	0,0460	0,0456	0,0476	0,0456	0,0490
	30	0,0468	0,0474	0,0436	0,0468	0,0394	0,0424	0,0440
	40	0,0464	0,0422	0,0410	0,0404	0,0406	0,0414	0,0456
	50	0,0434	0,0418	0,0464	0,0436	0,0422	0,0470	0,0480
	100	0,0420	0,0478	0,0434	0,0482	0,0434	0,0456	0,0544
	200	0,0570	0,0606	0,0570	0,0560	0,0670	0,0720	0,0656

Tabela 35 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0492	0,0470	0,0514	0,0548	0,0494	0,0464	0,0534
	30	0,0500	0,0524	0,0428	0,0482	0,0532	0,0532	0,0494
	40	0,0512	0,0492	0,0494	0,0526	0,0464	0,0508	0,0500
	50	0,0466	0,0456	0,0464	0,0488	0,0458	0,0588	0,0530
	100	0,0560	0,0462	0,0484	0,0510	0,0476	0,0486	0,0502
	200	0,0540	0,0508	0,0522	0,0474	0,0498	0,0522	0,0574
TKS	20	0,0526	0,0476	0,0514	0,0510	0,0582	0,0542	0,0566
	30	0,0524	0,0528	0,0490	0,0538	0,0490	0,0564	0,0546
	40	0,0522	0,0514	0,0450	0,0560	0,0560	0,0496	0,0568
	50	0,0428	0,0514	0,0526	0,0532	0,0518	0,0582	0,0570
	100	0,0480	0,0504	0,0558	0,0492	0,0472	0,0492	0,0606
	200	0,0602	0,0436	0,0500	0,0460	0,0490	0,0488	0,0566
TCM	20	0,0526	0,0496	0,0542	0,0480	0,0538	0,0544	0,0606
	30	0,0524	0,0502	0,0470	0,0556	0,0536	0,0578	0,0592
	40	0,0534	0,0522	0,0486	0,0554	0,0528	0,0512	0,0532
	50	0,0450	0,0484	0,0540	0,0510	0,0488	0,0536	0,0546
	100	0,0512	0,0508	0,0518	0,0452	0,0472	0,0508	0,0568
	200	0,0560	0,0446	0,0490	0,0492	0,0488	0,0526	0,0578
TAD	20	0,0558	0,0474	0,0570	0,0552	0,0584	0,0598	0,0672
	30	0,0530	0,0542	0,0520	0,0598	0,0570	0,0604	0,0632
	40	0,0508	0,0548	0,0498	0,0600	0,0500	0,0570	0,0568
	50	0,0446	0,0468	0,0490	0,0496	0,0540	0,0586	0,0646
	100	0,0530	0,0528	0,0522	0,0476	0,0504	0,0532	0,0592
	200	0,0542	0,0486	0,0498	0,0510	0,0492	0,0562	0,0618

Tabela 36 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ε						
		ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7
TRV	20	0,0510	0,0566	0,0468	0,0526	0,0528	0,0534	0,0622
	30	0,0446	0,0424	0,0532	0,0520	0,0542	0,0566	0,0604
	40	0,0520	0,0552	0,0470	0,0492	0,0504	0,0642	0,0644
	50	0,0506	0,0506	0,0486	0,0490	0,0536	0,0586	0,0768
	100	0,0574	0,0508	0,0456	0,0474	0,0588	0,0654	0,0920
	200	0,0492	0,0498	0,0494	0,0498	0,0570	0,0770	0,1238
TKS	20	0,0506	0,0622	0,0480	0,0536	0,0568	0,0648	0,0682
	30	0,0478	0,0456	0,0524	0,0518	0,0568	0,0604	0,0686
	40	0,0446	0,0510	0,0500	0,0526	0,0580	0,0698	0,0722
	50	0,0492	0,0510	0,0496	0,0508	0,0592	0,0630	0,0770
	100	0,0508	0,0460	0,0554	0,0538	0,0676	0,0646	0,0886
	200	0,0524	0,0498	0,0478	0,0482	0,0572	0,0798	0,1132
TCM	20	0,0496	0,0568	0,0462	0,0580	0,0596	0,0694	0,0720
	30	0,0466	0,0480	0,0522	0,0552	0,0556	0,0646	0,0688
	40	0,0492	0,0498	0,0534	0,0526	0,0592	0,0724	0,0752
	50	0,0522	0,0476	0,0524	0,0534	0,0532	0,0664	0,0816
	100	0,0556	0,0464	0,0588	0,0538	0,0684	0,0724	0,1008
	200	0,0498	0,0500	0,0460	0,0490	0,0614	0,0864	0,1254
TAD	20	0,0504	0,0594	0,0498	0,0604	0,0634	0,0762	0,0856
	30	0,0478	0,0488	0,0538	0,0550	0,0602	0,0744	0,0836
	40	0,0512	0,0548	0,0572	0,0570	0,0632	0,0812	0,0866
	50	0,0552	0,0492	0,0484	0,0532	0,0598	0,0702	0,0968
	100	0,0526	0,0496	0,0574	0,0532	0,0724	0,0818	0,1110
	200	0,0504	0,0478	0,0488	0,0502	0,0642	0,0938	0,1464

Tabela 37 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0490	0,0482	0,0510	0,0480	0,0584	0,0618	0,0818
	30	0,0506	0,0476	0,0508	0,0432	0,0550	0,0748	0,0940
	40	0,0530	0,0500	0,0452	0,0504	0,0574	0,0760	0,1088
	50	0,0496	0,0454	0,0570	0,0506	0,0550	0,0834	0,1152
	100	0,0480	0,0492	0,0514	0,0572	0,0686	0,1064	0,1894
	200	0,0554	0,0502	0,0484	0,0558	0,0934	0,1674	0,3168
TKS	20	0,0510	0,0464	0,0534	0,0520	0,0632	0,0752	0,0998
	30	0,0524	0,0460	0,0552	0,0546	0,0664	0,0804	0,1056
	40	0,0494	0,0560	0,0470	0,0568	0,0664	0,0860	0,1134
	50	0,0592	0,0526	0,0516	0,0534	0,0624	0,0798	0,1200
	100	0,0506	0,0532	0,0530	0,0552	0,0740	0,1088	0,1674
	200	0,0438	0,0494	0,0478	0,0568	0,0854	0,1542	0,2672
TCM	20	0,0474	0,0418	0,0500	0,0554	0,0650	0,0814	0,1060
	30	0,0522	0,0470	0,0556	0,0530	0,0652	0,0908	0,1124
	40	0,0502	0,0558	0,0458	0,0510	0,0696	0,0904	0,1300
	50	0,0538	0,0498	0,0540	0,0542	0,0652	0,0870	0,1346
	100	0,0476	0,0516	0,0556	0,0582	0,0732	0,1182	0,2010
	200	0,0452	0,0488	0,0452	0,0546	0,0924	0,1716	0,3200
TAD	20	0,0522	0,0470	0,0562	0,0648	0,0740	0,0962	0,1280
	30	0,0580	0,0498	0,0558	0,0516	0,0734	0,1082	0,1366
	40	0,0506	0,0552	0,0462	0,0578	0,0754	0,1070	0,1504
	50	0,0540	0,0488	0,0596	0,0612	0,0760	0,1036	0,1564
	100	0,0490	0,0514	0,0536	0,0626	0,0824	0,1328	0,2346
	200	0,0460	0,0508	0,0450	0,0610	0,1058	0,1952	0,3642

Tabela 38 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0530	0,0476	0,0482	0,0564	0,0628	0,0818	0,1142
	30	0,0550	0,0406	0,0480	0,0594	0,0708	0,0994	0,1500
	40	0,0460	0,0542	0,0490	0,0566	0,0728	0,1062	0,1840
	50	0,0546	0,0492	0,0472	0,0550	0,0828	0,1330	0,2190
	100	0,0458	0,0510	0,0514	0,0656	0,1110	0,2002	0,3614
	200	0,0468	0,0522	0,0574	0,0792	0,1728	0,3540	0,6132
TKS	20	0,0470	0,0536	0,0548	0,0636	0,0788	0,1038	0,1456
	30	0,0524	0,0428	0,0514	0,0650	0,0814	0,1036	0,1612
	40	0,0506	0,0584	0,0536	0,0592	0,0804	0,1200	0,1832
	50	0,0464	0,0496	0,0544	0,0682	0,0834	0,1266	0,2112
	100	0,0470	0,0476	0,0530	0,0684	0,1018	0,1886	0,3214
	200	0,0466	0,0546	0,0476	0,0726	0,1462	0,2928	0,5306
TCM	20	0,0470	0,0530	0,0556	0,0654	0,0782	0,1076	0,1546
	30	0,0474	0,0452	0,0496	0,0666	0,0894	0,1128	0,1764
	40	0,0544	0,0582	0,0520	0,0636	0,0844	0,1252	0,2176
	50	0,0488	0,0486	0,0534	0,0702	0,0942	0,1404	0,2456
	100	0,0480	0,0466	0,0544	0,0712	0,1124	0,2214	0,3876
	200	0,0466	0,0562	0,0528	0,0830	0,1668	0,3536	0,6336
TAD	20	0,0458	0,0522	0,0588	0,0738	0,0898	0,1286	0,1894
	30	0,0534	0,0454	0,0518	0,0724	0,0976	0,1422	0,2126
	40	0,0524	0,0574	0,0560	0,0722	0,0968	0,1526	0,2502
	50	0,0536	0,0508	0,0572	0,0728	0,1074	0,1718	0,2914
	100	0,0444	0,0478	0,0566	0,0758	0,1264	0,2526	0,4328
	200	0,0468	0,0536	0,0566	0,0916	0,1912	0,3922	0,6840

Nível nominal de $\alpha = 10\%$

Tabela 41 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0908	0,0874	0,0842	0,0840	0,0862	0,0966	0,0892
	30	0,0978	0,0852	0,0936	0,0888	0,0884	0,0868	0,0968
	40	0,0954	0,0894	0,0914	0,0912	0,0872	0,1008	0,0978
	50	0,0976	0,0926	0,0974	0,0976	0,0940	0,1020	0,0996
	100	0,1026	0,1020	0,1052	0,1042	0,1118	0,1130	0,1124
	200	0,1370	0,1362	0,1420	0,1398	0,1430	0,1560	0,1572
TKS	20	0,0960	0,0900	0,0920	0,0898	0,0948	0,0888	0,0884
	30	0,0934	0,0890	0,0916	0,0912	0,0928	0,0892	0,0870
	40	0,0888	0,0938	0,0860	0,0872	0,0910	0,0868	0,0920
	50	0,0902	0,0848	0,0964	0,0918	0,0934	0,0874	0,1044
	100	0,0960	0,1024	0,1058	0,0984	0,1028	0,1064	0,1092
	200	0,1186	0,1162	0,1188	0,1228	0,1304	0,1246	0,1336
TCM	20	0,0944	0,0920	0,0878	0,0886	0,0894	0,0890	0,0948
	30	0,0942	0,0922	0,0930	0,0874	0,0938	0,0888	0,0922
	40	0,0928	0,0972	0,0832	0,0874	0,0930	0,0940	0,0938
	50	0,0908	0,0846	0,0928	0,0926	0,0964	0,0946	0,1048
	100	0,0964	0,1024	0,1014	0,1006	0,0996	0,1080	0,1086
	200	0,1196	0,1202	0,1134	0,1198	0,1302	0,1374	0,1332
TAD	20	0,0980	0,0952	0,0908	0,0958	0,0940	0,0918	0,0976
	30	0,0926	0,0942	0,0910	0,0830	0,0896	0,0910	0,0956
	40	0,0914	0,0918	0,0840	0,0874	0,0838	0,0912	0,0944
	50	0,0924	0,0846	0,0916	0,0906	0,0958	0,0888	0,0944
	100	0,0944	0,0980	0,1010	0,0976	0,0940	0,1002	0,1060
	200	0,1212	0,1134	0,1160	0,1136	0,1260	0,1308	0,1316

Tabela 42 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0964	0,0958	0,1068	0,1068	0,1000	0,0998	0,1054
	30	0,0972	0,1066	0,0968	0,1002	0,1032	0,1064	0,1012
	40	0,0960	0,1008	0,0966	0,1034	0,0946	0,1012	0,0980
	50	0,0920	0,0962	0,0912	0,0952	0,0962	0,1092	0,1072
	100	0,1020	0,1042	0,0996	0,1022	0,0972	0,1006	0,1010
	200	0,1072	0,0924	0,0986	0,1004	0,0960	0,1006	0,1066
TKS	20	0,1014	0,0980	0,1006	0,1016	0,1082	0,1118	0,1096
	30	0,0982	0,1012	0,1002	0,1104	0,1002	0,1072	0,1088
	40	0,1034	0,1038	0,0972	0,1094	0,1036	0,1002	0,1094
	50	0,0900	0,1024	0,1032	0,1026	0,1014	0,1018	0,1080
	100	0,0984	0,0978	0,1112	0,0988	0,0966	0,0982	0,1072
	200	0,1078	0,0938	0,0968	0,0946	0,0996	0,0992	0,1164
TCM	20	0,1074	0,0982	0,1020	0,0992	0,1066	0,1040	0,1112
	30	0,1014	0,1034	0,1008	0,1118	0,1008	0,1102	0,1118
	40	0,1068	0,0990	0,0946	0,1056	0,1006	0,1046	0,1104
	50	0,0880	0,0956	0,1050	0,0990	0,0986	0,1052	0,1048
	100	0,0964	0,0988	0,1076	0,0960	0,0960	0,1016	0,1098
	200	0,1072	0,0924	0,0972	0,0998	0,0980	0,0992	0,1162
TAD	20	0,1098	0,0984	0,1092	0,1068	0,1066	0,1118	0,1234
	30	0,1072	0,1032	0,1014	0,1108	0,1070	0,1156	0,1196
	40	0,1052	0,1022	0,0984	0,1036	0,1048	0,1086	0,1130
	50	0,0912	0,0996	0,1034	0,1002	0,1006	0,1108	0,1146
	100	0,0982	0,1010	0,1050	0,1024	0,1002	0,1058	0,1150
	200	0,1046	0,0916	0,0990	0,0988	0,0990	0,1054	0,1196

Tabela 43 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1032	0,1042	0,0986	0,1052	0,1020	0,1088	0,1142
	30	0,0958	0,0950	0,1026	0,0976	0,1040	0,1112	0,1124
	40	0,1044	0,1078	0,1008	0,1026	0,0990	0,1180	0,1200
	50	0,0996	0,0952	0,1002	0,1004	0,1044	0,1074	0,1374
	100	0,1026	0,0990	0,1010	0,0952	0,1134	0,1282	0,1614
	200	0,0978	0,0970	0,0956	0,1038	0,1120	0,1444	0,2048
TKS	20	0,1012	0,1104	0,1002	0,1106	0,1148	0,1168	0,1262
	30	0,0960	0,0998	0,0984	0,1038	0,1052	0,1144	0,1270
	40	0,0934	0,1000	0,1040	0,1050	0,1162	0,1284	0,1282
	50	0,1042	0,0996	0,1030	0,1016	0,1098	0,1132	0,1354
	100	0,1042	0,0990	0,1036	0,1038	0,1254	0,1276	0,1596
	200	0,1014	0,1034	0,0934	0,0958	0,1112	0,1338	0,1842
TCM	20	0,1038	0,1076	0,0976	0,1134	0,1126	0,1194	0,1334
	30	0,1018	0,1008	0,0988	0,1068	0,1064	0,1228	0,1318
	40	0,0920	0,1014	0,1056	0,1010	0,1160	0,1332	0,1312
	50	0,1036	0,0988	0,1090	0,1054	0,1080	0,1254	0,1368
	100	0,1048	0,0940	0,1104	0,1068	0,1242	0,1326	0,1652
	200	0,1012	0,0994	0,0946	0,1026	0,1150	0,1476	0,2066
TAD	20	0,0972	0,1084	0,1018	0,1106	0,1200	0,1328	0,1514
	30	0,1038	0,1016	0,1024	0,1056	0,1136	0,1318	0,1468
	40	0,0962	0,1066	0,1084	0,1086	0,1198	0,1408	0,1490
	50	0,1068	0,1012	0,1094	0,1080	0,1172	0,1332	0,1588
	100	0,1100	0,0970	0,1082	0,1092	0,1302	0,1436	0,1816
	200	0,0982	0,1016	0,0968	0,1032	0,1182	0,1574	0,2234

Tabela 44 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1028	0,0916	0,1030	0,1050	0,1064	0,1222	0,1440
	30	0,1004	0,0942	0,0990	0,0892	0,1124	0,1336	0,1664
	40	0,1032	0,1050	0,0932	0,1030	0,1124	0,1386	0,1882
	50	0,1010	0,0936	0,1080	0,1042	0,1158	0,1498	0,1966
	100	0,1030	0,1032	0,0962	0,1112	0,1320	0,1774	0,2904
	200	0,1036	0,0984	0,1000	0,1118	0,1606	0,2636	0,4352
TKS	20	0,1046	0,0922	0,1058	0,1048	0,1186	0,1432	0,1706
	30	0,1026	0,0976	0,1044	0,1052	0,1254	0,1456	0,1716
	40	0,1032	0,1172	0,0942	0,1040	0,1250	0,1536	0,1854
	50	0,1144	0,1000	0,1022	0,1056	0,1172	0,1428	0,2000
	100	0,0980	0,0982	0,0976	0,1074	0,1334	0,1768	0,2616
	200	0,0914	0,1012	0,0986	0,1134	0,1532	0,2382	0,3960
TCM	20	0,1064	0,0900	0,1036	0,1032	0,1178	0,1446	0,1794
	30	0,1036	0,0970	0,1056	0,1058	0,1270	0,1508	0,1860
	40	0,1012	0,1198	0,0942	0,1060	0,1270	0,1606	0,1988
	50	0,1148	0,0998	0,1082	0,1032	0,1200	0,1510	0,2134
	100	0,0936	0,0996	0,1046	0,1114	0,1348	0,1880	0,3008
	200	0,0948	0,0972	0,0936	0,1062	0,1574	0,2696	0,4442
TAD	20	0,1002	0,0936	0,1068	0,1152	0,1318	0,1642	0,2090
	30	0,1032	0,1024	0,1092	0,1058	0,1354	0,1754	0,2182
	40	0,1018	0,1170	0,0944	0,1098	0,1370	0,1766	0,2364
	50	0,1106	0,0948	0,1084	0,1102	0,1352	0,1708	0,2454
	100	0,0950	0,1032	0,1028	0,1162	0,1448	0,2056	0,3342
	200	0,0932	0,0980	0,0972	0,1140	0,1704	0,2904	0,4850

Tabela 45 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0996	0,0968	0,0978	0,1122	0,1148	0,1540	0,1904
	30	0,1046	0,0908	0,1004	0,1108	0,1296	0,1798	0,2358
	40	0,0948	0,1012	0,0938	0,1066	0,1312	0,1838	0,2718
	50	0,1010	0,1022	0,0962	0,1108	0,1496	0,2142	0,3160
	100	0,0942	0,0954	0,1010	0,1220	0,1842	0,3008	0,4822
	200	0,0944	0,0978	0,1074	0,1488	0,2654	0,4680	0,7160
TKS	20	0,0986	0,1020	0,1044	0,1168	0,1372	0,1710	0,2280
	30	0,1014	0,0920	0,1004	0,1202	0,1366	0,1780	0,2496
	40	0,1008	0,1146	0,0994	0,1132	0,1428	0,1914	0,2802
	50	0,1004	0,0968	0,1046	0,1254	0,1510	0,2052	0,3132
	100	0,0958	0,0966	0,1030	0,1228	0,1646	0,2868	0,4444
	200	0,0950	0,1028	0,0990	0,1394	0,2332	0,4108	0,6550
TCM	20	0,0928	0,1034	0,1044	0,1180	0,1392	0,1802	0,2410
	30	0,1018	0,0890	0,1008	0,1214	0,1472	0,1912	0,2682
	40	0,0998	0,1136	0,1028	0,1174	0,1470	0,2108	0,3126
	50	0,1002	0,0980	0,1098	0,1254	0,1586	0,2268	0,3512
	100	0,0968	0,0946	0,1070	0,1278	0,1866	0,3186	0,5088
	200	0,0966	0,1084	0,1066	0,1426	0,2640	0,4742	0,7332
TAD	20	0,0950	0,1066	0,1064	0,1318	0,1586	0,2086	0,2786
	30	0,1000	0,0904	0,1048	0,1336	0,1632	0,2268	0,3100
	40	0,1020	0,1104	0,1070	0,1272	0,1608	0,2386	0,3524
	50	0,0994	0,1044	0,1092	0,1344	0,1758	0,2564	0,3874
	100	0,0912	0,0994	0,1096	0,1390	0,2044	0,3502	0,5512
	200	0,0918	0,1096	0,1092	0,1566	0,2940	0,5186	0,7790

Tabela 46 – Taxas de rejeição de H_0 : os dados possuem distribuição Watson complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD, para $\kappa = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1066	0,1032	0,1150	0,1578	0,2948	0,4620	0,6312
	30	0,0978	0,0984	0,1270	0,1822	0,3618	0,5678	0,7788
	40	0,1010	0,1018	0,1252	0,2210	0,4274	0,6850	0,8708
	50	0,0970	0,1012	0,1282	0,2598	0,4986	0,7508	0,9168
	100	0,1010	0,1046	0,1580	0,3902	0,7280	0,9496	0,9928
	200	0,1006	0,1130	0,2234	0,5950	0,9380	0,9988	1,0000
TKS	20	0,1030	0,1048	0,1306	0,1844	0,3190	0,4804	0,6480
	30	0,0944	0,0932	0,1296	0,2136	0,3776	0,5786	0,7758
	40	0,0964	0,1028	0,1430	0,2224	0,4162	0,6642	0,8584
	50	0,1042	0,0966	0,1292	0,2574	0,4850	0,7324	0,9010
	100	0,1076	0,1102	0,1576	0,3546	0,6814	0,9250	0,9938
	200	0,1002	0,1092	0,2010	0,5340	0,9148	0,9966	1,0000
TCM	20	0,0980	0,1080	0,1310	0,1930	0,3492	0,5278	0,7164
	30	0,1004	0,0978	0,1334	0,2276	0,4158	0,6352	0,8364
	40	0,1008	0,1028	0,1434	0,2416	0,4626	0,7356	0,9086
	50	0,1052	0,1038	0,1324	0,2782	0,5354	0,8032	0,9454
	100	0,1080	0,1118	0,1686	0,3996	0,7608	0,9618	0,9984
	200	0,0980	0,1142	0,2200	0,6060	0,9530	0,9996	1,0000
TAD	20	0,1008	0,1176	0,1484	0,2344	0,4102	0,6058	0,7814
	30	0,0992	0,1054	0,1542	0,2594	0,4698	0,6918	0,8798
	40	0,1004	0,1100	0,1516	0,2750	0,5250	0,7930	0,9380
	50	0,1020	0,1042	0,1518	0,3186	0,5960	0,8502	0,9644
	100	0,1086	0,1236	0,1886	0,4414	0,8040	0,9778	0,9986
	200	0,0964	0,1188	0,2412	0,6464	0,9690	0,9998	1,0000

APÊNDICE B – SIMULAÇÕES DOS TESTES DE BONDADE DE AJUSTE - BINGHAM COMPLEXA

NÍVEL NOMINAL - PARÂMETROS CONHECIDOS - BINGHAM COMPLEXA

Nas Tabelas 21 e 22 estão as simulações para os tamanhos de 1% e 5% quando os parâmetros são conhecidos, respectivamente.

Tabela 49 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.

Teste	n	σ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	0,0436	0,0096	0,0098	0,0096	0,0084	0,0104	0,0110	0,0094	0,0094	0,0120	0,0068
	30	1,0000	0,0898	0,0102	0,0074	0,0086	0,0090	0,0086	0,0102	0,0106	0,0112	0,0098	0,0096
	40	1,0000	0,1344	0,0078	0,0094	0,0098	0,0126	0,0106	0,0106	0,0094	0,0088	0,0112	0,0106
	50	1,0000	0,1888	0,0120	0,0076	0,0112	0,0106	0,0106	0,0098	0,0082	0,0094	0,0096	0,0100
	100	1,0000	0,5024	0,0152	0,0104	0,0124	0,0088	0,0102	0,0108	0,0106	0,0082	0,0112	0,0080
TKS	200	1,0000	0,8990	0,0218	0,0098	0,0098	0,0100	0,0068	0,0118	0,0072	0,0088	0,0094	0,0068
	20	1,0000	0,0538	0,0100	0,0110	0,0076	0,0096	0,0090	0,0098	0,0104	0,0092	0,0104	0,0070
	30	1,0000	0,1042	0,0114	0,0080	0,0106	0,0096	0,0092	0,0104	0,0110	0,0110	0,0104	0,0092
	40	1,0000	0,1520	0,0092	0,0088	0,0104	0,0096	0,0078	0,0100	0,0092	0,0082	0,0102	0,0100
	50	1,0000	0,2140	0,0126	0,0082	0,0098	0,0108	0,0090	0,0114	0,0088	0,0112	0,0110	0,0124
TCM	100	1,0000	0,5610	0,0134	0,0106	0,0090	0,0084	0,0072	0,0074	0,0094	0,0066	0,0094	0,0074
	200	1,0000	0,9592	0,0160	0,0080	0,0096	0,0104	0,0090	0,0102	0,0090	0,0094	0,0078	0,0070
	20	1,0000	0,0646	0,0104	0,0102	0,0098	0,0100	0,0094	0,0102	0,0108	0,0086	0,0108	0,0074
	30	1,0000	0,1198	0,0126	0,0072	0,0076	0,0090	0,0096	0,0092	0,0100	0,0098	0,0094	0,0090
	40	1,0000	0,1872	0,0084	0,0086	0,0098	0,0122	0,0106	0,0092	0,0074	0,0088	0,0108	0,0086
TAD	50	1,0000	0,2616	0,0126	0,0078	0,0098	0,0112	0,0092	0,0102	0,0086	0,0096	0,0100	0,0120
	100	1,0000	0,6520	0,0178	0,0110	0,0116	0,0096	0,0102	0,0108	0,0108	0,0090	0,0108	0,0074
	200	1,0000	0,9734	0,0210	0,0084	0,0106	0,0094	0,0086	0,0122	0,0090	0,0092	0,0082	0,0072
	20	1,0000	0,0570	0,0102	0,0090	0,0094	0,0094	0,0112	0,0092	0,0108	0,0082	0,0120	0,0080
	30	1,0000	0,1138	0,0126	0,0074	0,0082	0,0102	0,0090	0,0104	0,0112	0,0110	0,0098	0,0086
TAD	40	1,0000	0,1844	0,0088	0,0088	0,0098	0,0122	0,0108	0,0086	0,0084	0,0096	0,0108	0,0098
	50	1,0000	0,2670	0,0124	0,0082	0,0104	0,0112	0,0100	0,0096	0,0072	0,0096	0,0110	0,0118
	100	1,0000	0,7062	0,0174	0,0098	0,0130	0,0092	0,0090	0,0112	0,0118	0,0084	0,0118	0,0084
	200	1,0000	0,9908	0,0240	0,0096	0,0102	0,0098	0,0094	0,0132	0,0076	0,0094	0,0076	0,0076

Tabela 50 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.

Teste	n	σ											
		1	3	5	8	10	20	30	40	50	100	1000	5000
TRV	20	1,0000	0,2646	0,0996	0,1014	0,1022	0,0970	0,0950	0,1010	0,1000	0,0968	0,1016	0,0954
	30	1,0000	0,3878	0,0972	0,0934	0,0966	0,0974	0,0960	0,0936	0,1024	0,1062	0,0980	0,1062
	40	1,0000	0,4932	0,0970	0,1010	0,1030	0,1024	0,0998	0,0984	0,1060	0,1052	0,1046	0,0974
	50	1,0000	0,5840	0,1050	0,1014	0,1058	0,1056	0,0990	0,0900	0,0954	0,1000	0,0922	0,1070
	100	1,0000	0,8664	0,1224	0,0970	0,0964	0,0960	0,0936	0,1032	0,1022	0,1042	0,1012	0,0924
	200	1,0000	0,9924	0,1584	0,0996	0,1034	0,0970	0,0960	0,1002	0,0972	0,0966	0,0968	0,1064
TKS	20	1,0000	0,2886	0,1050	0,0966	0,0958	0,1006	0,0978	0,0964	0,0942	0,0978	0,1014	0,0976
	30	1,0000	0,4358	0,1034	0,1000	0,0936	0,0978	0,0964	0,1006	0,1022	0,1066	0,1058	0,1060
	40	1,0000	0,5358	0,1048	0,0986	0,1002	0,1064	0,0960	0,1012	0,1062	0,1018	0,1044	0,1010
	50	1,0000	0,6426	0,1072	0,1026	0,0998	0,1008	0,1058	0,0946	0,0896	0,0968	0,0910	0,1006
	100	1,0000	0,9202	0,1082	0,0952	0,0940	0,0900	0,0830	0,0956	0,0852	0,0862	0,0920	0,0964
	200	1,0000	0,9990	0,1230	0,0998	0,0942	0,0858	0,0962	0,0906	0,0910	0,0914	0,0958	0,0956
TCM	20	1,0000	0,3214	0,1026	0,0984	0,0994	0,0990	0,0988	0,0970	0,0928	0,0982	0,1016	0,0940
	30	1,0000	0,4742	0,1012	0,0960	0,0996	0,0978	0,0946	0,0962	0,1006	0,1086	0,0982	0,1078
	40	1,0000	0,5862	0,1028	0,0994	0,1014	0,1068	0,0980	0,1000	0,1066	0,1048	0,1036	0,1000
	50	1,0000	0,6900	0,1058	0,1030	0,0986	0,1014	0,1016	0,0922	0,0920	0,0988	0,0936	0,1008
	100	1,0000	0,9416	0,1192	0,1018	0,1024	0,0978	0,0938	0,1088	0,0974	0,0952	0,0984	0,1008
	200	1,0000	0,9994	0,1472	0,1026	0,1024	0,0954	0,0940	0,0976	0,0972	0,0994	0,1016	0,0994
TAD	20	1,0000	0,3140	0,0984	0,0946	0,0974	0,0994	0,1014	0,0968	0,0956	0,0976	0,1008	0,0956
	30	1,0000	0,4758	0,0996	0,0892	0,0998	0,0980	0,0912	0,0924	0,1036	0,1082	0,1004	0,1040
	40	1,0000	0,6086	0,1024	0,0972	0,1000	0,1046	0,0912	0,0994	0,1076	0,1062	0,1032	0,0992
	50	1,0000	0,7276	0,1036	0,1014	0,1004	0,1026	0,1018	0,0898	0,0930	0,1026	0,0936	0,1026
	100	1,0000	0,9690	0,1252	0,1014	0,1014	0,0998	0,0976	0,1096	0,0954	0,0958	0,1016	0,0992
	200	1,0000	0,9998	0,1704	0,1010	0,1016	0,0974	0,0948	0,1000	0,0958	0,0986	0,0984	0,1018

NÍVEL NOMINAL - PARÂMETROS ESTIMADOS - BINGHAM COMPLEXA

Nas Tabelas 23, 24 e 25 estão as simulações para os tamanhos de 1%, 5% e 10% quando os parâmetros são desconhecidos, respectivamente.

Tabela 51 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 1%.

Teste	n	σ								
		8	10	20	30	40	50	100	1000	5000
TRV	20	0,0182	0,0216	0,0216	0,0230	0,0224	0,0224	0,0222	0,0186	0,0176
	30	0,0146	0,0150	0,0178	0,0190	0,0162	0,0192	0,0128	0,0176	0,0160
	40	0,0122	0,0138	0,0122	0,0166	0,0136	0,0114	0,0144	0,0148	0,0124
	50	0,0116	0,0120	0,0122	0,0136	0,0118	0,0164	0,0144	0,0162	0,0126
	100	0,0102	0,0110	0,0100	0,0118	0,0104	0,0110	0,0162	0,0100	0,0082
	200	0,0112	0,0098	0,0082	0,0098	0,0120	0,0116	0,0132	0,0102	0,0102
TKS	20	0,0106	0,0108	0,0102	0,0130	0,0094	0,0122	0,0124	0,0122	0,0120
	30	0,0096	0,0108	0,0100	0,0102	0,0092	0,0110	0,0094	0,0082	0,0096
	40	0,0086	0,0110	0,0090	0,0106	0,0106	0,0104	0,0114	0,0092	0,0080
	50	0,0082	0,0108	0,0106	0,0086	0,0094	0,0104	0,0074	0,0108	0,0118
	100	0,0088	0,0080	0,0094	0,0096	0,0094	0,0082	0,0096	0,0118	0,0076
	200	0,0112	0,0104	0,0112	0,0094	0,0102	0,0114	0,0098	0,0110	0,0118
TCM	20	0,0114	0,0140	0,0112	0,0146	0,0124	0,0128	0,0148	0,0116	0,0114
	30	0,0104	0,0112	0,0140	0,0132	0,0106	0,0120	0,0066	0,0082	0,0112
	40	0,0090	0,0134	0,0094	0,0112	0,0126	0,0112	0,0128	0,0088	0,0086
	50	0,0086	0,0104	0,0096	0,0106	0,0088	0,0100	0,0122	0,0108	0,0138
	100	0,0086	0,0084	0,0090	0,0104	0,0104	0,0086	0,0132	0,0118	0,0078
	200	0,0098	0,0106	0,0120	0,0120	0,0106	0,0106	0,0114	0,0120	0,0100
TAD	20	0,0048	0,0086	0,0040	0,0070	0,0050	0,0066	0,0068	0,0062	0,0072
	30	0,0064	0,0070	0,0090	0,0074	0,0070	0,0088	0,0054	0,0062	0,0080
	40	0,0082	0,0090	0,0062	0,0086	0,0092	0,0074	0,0086	0,0066	0,0064
	50	0,0062	0,0076	0,0074	0,0076	0,0070	0,0100	0,0098	0,0088	0,0094
	100	0,0082	0,0086	0,0080	0,0088	0,0102	0,0084	0,0124	0,0102	0,0072
	200	0,0102	0,0112	0,0106	0,0108	0,0102	0,0114	0,0124	0,0104	0,0100

Tabela 52 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 5%.

Teste	n	σ								
		8	10	20	30	40	50	100	1000	5000
TRV	20	0,0774	0,0786	0,0880	0,0838	0,0892	0,0862	0,0918	0,0812	0,0732
	30	0,0652	0,0638	0,0702	0,0688	0,0710	0,0726	0,0678	0,0680	0,0642
	40	0,0640	0,0614	0,0588	0,0678	0,0630	0,0652	0,0666	0,0668	0,0590
	50	0,0584	0,0586	0,0574	0,0582	0,0616	0,0640	0,0594	0,0558	0,0570
	100	0,0596	0,0522	0,0538	0,0566	0,0542	0,0546	0,0604	0,0502	0,0422
	200	0,0500	0,0488	0,0514	0,0458	0,0572	0,0552	0,0586	0,0550	0,0556
TKS	20	0,0508	0,0524	0,0532	0,0528	0,0600	0,0592	0,0562	0,0588	0,0570
	30	0,0490	0,0474	0,0558	0,0558	0,0524	0,0522	0,0500	0,0506	0,0512
	40	0,0454	0,0534	0,0544	0,0512	0,0488	0,0526	0,0536	0,0546	0,0468
	50	0,0492	0,0500	0,0508	0,0484	0,0504	0,0480	0,0520	0,0498	0,0504
	100	0,0474	0,0492	0,0450	0,0518	0,0530	0,0490	0,0532	0,0492	0,0468
	200	0,0518	0,0464	0,0512	0,0512	0,0508	0,0544	0,0570	0,0500	0,0516
TCM	20	0,0560	0,0540	0,0584	0,0570	0,0622	0,0640	0,0590	0,0584	0,0588
	30	0,0470	0,0496	0,0530	0,0582	0,0558	0,0546	0,0490	0,0566	0,0524
	40	0,0462	0,0508	0,0566	0,0550	0,0522	0,0590	0,0538	0,0538	0,0430
	50	0,0494	0,0484	0,0490	0,0504	0,0508	0,0468	0,0542	0,0514	0,0518
	100	0,0520	0,0512	0,0482	0,0484	0,0542	0,0470	0,0562	0,0482	0,0446
	200	0,0512	0,0502	0,0494	0,0470	0,0520	0,0526	0,0562	0,0534	0,0516
TAD	20	0,0376	0,0426	0,0420	0,0442	0,0432	0,0450	0,0444	0,0424	0,0452
	30	0,0374	0,0416	0,0456	0,0482	0,0434	0,0440	0,0352	0,0450	0,0462
	40	0,0418	0,0444	0,0470	0,0484	0,0426	0,0510	0,0492	0,0482	0,0392
	50	0,0434	0,0458	0,0418	0,0436	0,0468	0,0432	0,0472	0,0478	0,0478
	100	0,0528	0,0472	0,0448	0,0484	0,0524	0,0464	0,0544	0,0454	0,0422
	200	0,0500	0,0486	0,0490	0,0472	0,0498	0,0524	0,0546	0,0538	0,0512

Tabela 53 – Taxas de rejeição de H_0 : os dados possuem distribuição Binghma complexa, utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, diferentes valores de σ , diversos tamanhos amostrais e utilizando o nível nominal de 10%.

Teste	n	σ								
		8	10	20	30	40	50	100	1000	5000
TRV	20	0,1398	0,1360	0,1528	0,1490	0,1570	0,1546	0,1576	0,1410	0,1356
	30	0,1218	0,1228	0,1302	0,1306	0,1324	0,1382	0,1278	0,1222	0,1134
	40	0,1204	0,1136	0,1186	0,1250	0,1228	0,1242	0,1272	0,1260	0,1128
	50	0,1080	0,1108	0,1148	0,1096	0,1206	0,1118	0,1140	0,1178	0,1158
	100	0,1128	0,1004	0,1038	0,1102	0,1030	0,1110	0,1082	0,1024	0,0920
	200	0,1052	0,0976	0,1006	0,0988	0,1080	0,1094	0,1090	0,1070	0,1096
TKS	20	0,1084	0,1036	0,1086	0,1050	0,1122	0,1104	0,1106	0,1128	0,1134
	30	0,0994	0,0932	0,1048	0,1076	0,1018	0,1034	0,1010	0,1030	0,1032
	40	0,0988	0,1040	0,1040	0,1068	0,0998	0,1042	0,1062	0,1074	0,0966
	50	0,1020	0,1066	0,1092	0,0966	0,1008	0,0996	0,1020	0,0996	0,0976
	100	0,0994	0,0950	0,0954	0,1036	0,1008	0,0954	0,0982	0,0946	0,0940
	200	0,1026	0,0958	0,0988	0,0990	0,0988	0,1030	0,1120	0,1108	0,1028
TCM	20	0,1020	0,1074	0,1122	0,1114	0,1192	0,1164	0,1160	0,1170	0,1146
	30	0,0970	0,0964	0,1038	0,1084	0,1094	0,1086	0,1112	0,1060	0,1074
	40	0,1012	0,1046	0,1080	0,1116	0,0988	0,1058	0,1082	0,1134	0,0932
	50	0,0996	0,1066	0,1042	0,0992	0,1020	0,0970	0,1068	0,1012	0,0988
	100	0,1054	0,0972	0,1006	0,0980	0,1072	0,0998	0,1090	0,0970	0,0930
	200	0,0982	0,1004	0,0984	0,1004	0,0974	0,1020	0,1134	0,1118	0,1052
TAD	20	0,0894	0,0910	0,0920	0,0922	0,0958	0,0918	0,0922	0,0932	0,0962
	30	0,0832	0,0902	0,0924	0,0958	0,0988	0,0924	0,0920	0,0930	0,0970
	40	0,0924	0,0944	0,1006	0,1032	0,0882	0,0980	0,0990	0,1000	0,0870
	50	0,0878	0,0950	0,0928	0,0942	0,0944	0,0910	0,0956	0,1004	0,0908
	100	0,1014	0,0922	0,0968	0,0970	0,1010	0,0956	0,1012	0,0932	0,0852
	200	0,1008	0,0942	0,0940	0,0968	0,0996	0,1018	0,1064	0,1094	0,1028

ANÁLISE DO PODER - BINGHAM COMPLEXA

Nível nominal de $\alpha = 1\%$

Tabela 54 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0152	0,0142	0,0136	0,0138	0,0148	0,0110	0,0116
	30	0,0116	0,0110	0,0112	0,0126	0,0106	0,0094	0,0138
	40	0,0140	0,0100	0,0104	0,0118	0,0108	0,0126	0,0116
	50	0,0130	0,0110	0,0088	0,0098	0,0106	0,0108	0,0112
	100	0,0132	0,0080	0,0116	0,0112	0,0120	0,0104	0,0118
	200	0,0134	0,0124	0,0188	0,0160	0,0208	0,0216	0,0228
TKS	20	0,0100	0,0098	0,0116	0,0096	0,0080	0,0082	0,0080
	30	0,0088	0,0106	0,0082	0,0110	0,0080	0,0100	0,0112
	40	0,0084	0,0098	0,0084	0,0078	0,0108	0,0116	0,0106
	50	0,0078	0,0084	0,0114	0,0106	0,0112	0,0118	0,0068
	100	0,0116	0,0092	0,0124	0,0122	0,0126	0,0106	0,0136
	200	0,0144	0,0118	0,0170	0,0172	0,0160	0,0180	0,0250
TCM	20	0,0112	0,0082	0,0112	0,0084	0,0088	0,0092	0,0088
	30	0,0076	0,0108	0,0106	0,0102	0,0096	0,0094	0,0112
	40	0,0102	0,0070	0,0092	0,0088	0,0126	0,0116	0,0116
	50	0,0102	0,0094	0,0114	0,0130	0,0122	0,0124	0,0092
	100	0,0110	0,0114	0,0124	0,0102	0,0128	0,0114	0,0142
	200	0,0148	0,0106	0,0182	0,0154	0,0182	0,0200	0,0228
TAD	20	0,0062	0,0052	0,0068	0,0060	0,0058	0,0052	0,0064
	30	0,0064	0,0078	0,0088	0,0076	0,0082	0,0090	0,0090
	40	0,0098	0,0078	0,0098	0,0074	0,0104	0,0114	0,0110
	50	0,0124	0,0098	0,0098	0,0108	0,0128	0,0118	0,0096
	100	0,0136	0,0100	0,0128	0,0108	0,0134	0,0142	0,0146
	200	0,0160	0,0138	0,0206	0,0156	0,0188	0,0246	0,0260

Tabela 55 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0126	0,0164	0,0148	0,0142	0,0138	0,0124	0,0118
	30	0,0110	0,0120	0,0132	0,0106	0,0122	0,0118	0,0128
	40	0,0124	0,0118	0,0134	0,0110	0,0110	0,0112	0,0116
	50	0,0118	0,0098	0,0124	0,0104	0,0134	0,0126	0,0142
	100	0,0128	0,0136	0,0148	0,0120	0,0132	0,0164	0,0166
	200	0,0208	0,0182	0,0184	0,0214	0,0260	0,0394	0,0446
TKS	20	0,0124	0,0072	0,0106	0,0088	0,0086	0,0090	0,0092
	30	0,0082	0,0094	0,0078	0,0094	0,0096	0,0080	0,0128
	40	0,0092	0,0092	0,0096	0,0138	0,0114	0,0122	0,0140
	50	0,0094	0,0092	0,0096	0,0116	0,0106	0,0134	0,0172
	100	0,0116	0,0136	0,0104	0,0112	0,0150	0,0126	0,0186
	200	0,0152	0,0176	0,0186	0,0180	0,0258	0,0270	0,0340
TCM	20	0,0102	0,0102	0,0104	0,0080	0,0094	0,0114	0,0086
	30	0,0086	0,0098	0,0078	0,0094	0,0106	0,0088	0,0124
	40	0,0104	0,0082	0,0114	0,0120	0,0120	0,0114	0,0122
	50	0,0094	0,0108	0,0122	0,0106	0,0098	0,0148	0,0160
	100	0,0124	0,0158	0,0132	0,0130	0,0170	0,0158	0,0194
	200	0,0178	0,0160	0,0190	0,0184	0,0268	0,0356	0,0414
TAD	20	0,0072	0,0070	0,0064	0,0058	0,0054	0,0064	0,0050
	30	0,0080	0,0072	0,0066	0,0098	0,0108	0,0094	0,0112
	40	0,0094	0,0092	0,0128	0,0112	0,0114	0,0118	0,0124
	50	0,0098	0,0110	0,0134	0,0116	0,0106	0,0170	0,0192
	100	0,0150	0,0144	0,0148	0,0138	0,0162	0,0188	0,0218
	200	0,0184	0,0194	0,0192	0,0234	0,0290	0,0390	0,0490

Tabela 56 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0154	0,0120	0,0146	0,0138	0,0122	0,0102	0,0054
	30	0,0138	0,0132	0,0114	0,0098	0,0116	0,0108	0,0138
	40	0,0124	0,0096	0,0118	0,0104	0,0132	0,0160	0,0176
	50	0,0124	0,0098	0,0110	0,0134	0,0138	0,0182	0,0266
	100	0,0102	0,0142	0,0164	0,0234	0,0334	0,0544	0,0926
	200	0,0198	0,0232	0,0324	0,0528	0,0966	0,1560	0,2504
TKS	20	0,0064	0,0094	0,0086	0,0080	0,0088	0,0076	0,0086
	30	0,0096	0,0082	0,0124	0,0088	0,0122	0,0142	0,0162
	40	0,0106	0,0102	0,0092	0,0128	0,0130	0,0164	0,0236
	50	0,0118	0,0132	0,0116	0,0130	0,0184	0,0234	0,0320
	100	0,0120	0,0140	0,0158	0,0224	0,0286	0,0448	0,0718
	200	0,0190	0,0210	0,0282	0,0398	0,0732	0,1086	0,1836
TCM	20	0,0088	0,0088	0,0090	0,0088	0,0090	0,0084	0,0076
	30	0,0100	0,0084	0,0106	0,0084	0,0136	0,0140	0,0172
	40	0,0094	0,0102	0,0064	0,0112	0,0154	0,0188	0,0260
	50	0,0120	0,0124	0,0122	0,0144	0,0182	0,0228	0,0358
	100	0,0118	0,0164	0,0184	0,0218	0,0368	0,0518	0,0954
	200	0,0212	0,0214	0,0312	0,0494	0,0922	0,1528	0,2434
TAD	20	0,0036	0,0058	0,0052	0,0068	0,0090	0,0078	0,0092
	30	0,0106	0,0104	0,0100	0,0088	0,0126	0,0168	0,0218
	40	0,0084	0,0092	0,0084	0,0110	0,0166	0,0210	0,0334
	50	0,0104	0,0108	0,0134	0,0142	0,0216	0,0292	0,0422
	100	0,0130	0,0174	0,0202	0,0260	0,0422	0,0624	0,1090
	200	0,0214	0,0240	0,0372	0,0558	0,1058	0,1760	0,2780

Tabela 57 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0168	0,0138	0,0130	0,0096	0,0086	0,0094	0,0058
	30	0,0120	0,0116	0,0114	0,0096	0,0122	0,0172	0,0240
	40	0,0116	0,0136	0,0124	0,0156	0,0190	0,0280	0,0406
	50	0,0114	0,0124	0,0130	0,0172	0,0296	0,0432	0,0634
	100	0,0156	0,0162	0,0222	0,0396	0,0814	0,1366	0,2110
	200	0,0196	0,0284	0,0550	0,1096	0,2330	0,3986	0,5696
TKS	20	0,0076	0,0094	0,0094	0,0076	0,0116	0,0098	0,0124
	30	0,0084	0,0100	0,0104	0,0128	0,0160	0,0196	0,0302
	40	0,0094	0,0090	0,0132	0,0180	0,0244	0,0338	0,0498
	50	0,0114	0,0120	0,0132	0,0220	0,0284	0,0474	0,0648
	100	0,0138	0,0148	0,0212	0,0364	0,0684	0,1150	0,1666
	200	0,0156	0,0258	0,0422	0,0862	0,1758	0,2912	0,4322
TCM	20	0,0100	0,0108	0,0096	0,0096	0,0100	0,0094	0,0114
	30	0,0096	0,0086	0,0090	0,0124	0,0158	0,0236	0,0324
	40	0,0094	0,0098	0,0144	0,0186	0,0244	0,0352	0,0590
	50	0,0120	0,0110	0,0158	0,0226	0,0306	0,0524	0,0784
	100	0,0146	0,0178	0,0202	0,0456	0,0872	0,1496	0,2256
	200	0,0192	0,0296	0,0474	0,1100	0,2300	0,3970	0,5914
TAD	20	0,0050	0,0054	0,0074	0,0080	0,0104	0,0110	0,0142
	30	0,0074	0,0070	0,0094	0,0106	0,0202	0,0272	0,0428
	40	0,0086	0,0096	0,0140	0,0242	0,0310	0,0432	0,0720
	50	0,0098	0,0114	0,0166	0,0252	0,0384	0,0662	0,0970
	100	0,0168	0,0192	0,0222	0,0520	0,1008	0,1782	0,2658
	200	0,0196	0,0302	0,0560	0,1260	0,2630	0,4436	0,6436

Tabela 58 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0164	0,0138	0,0084	0,0086	0,0090	0,0120	0,0140
	30	0,0142	0,0112	0,0122	0,0108	0,0158	0,0272	0,0362
	40	0,0126	0,0126	0,0142	0,0188	0,0282	0,0466	0,0684
	50	0,0112	0,0138	0,0106	0,0272	0,0430	0,0762	0,1034
	100	0,0146	0,0202	0,0334	0,0792	0,1516	0,2648	0,3856
	200	0,0216	0,0352	0,0886	0,2156	0,4230	0,6440	0,7964
TKS	20	0,0110	0,0104	0,0080	0,0106	0,0108	0,0172	0,0192
	30	0,0112	0,0104	0,0108	0,0146	0,0222	0,0296	0,0472
	40	0,0104	0,0130	0,0146	0,0226	0,0348	0,0486	0,0730
	50	0,0100	0,0118	0,0170	0,0256	0,0466	0,0760	0,1038
	100	0,0134	0,0196	0,0338	0,0670	0,1168	0,2076	0,3046
	200	0,0180	0,0274	0,0632	0,1532	0,2958	0,5036	0,6764
TCM	20	0,0100	0,0098	0,0070	0,0094	0,0110	0,0164	0,0200
	30	0,0096	0,0094	0,0136	0,0136	0,0214	0,0416	0,0518
	40	0,0104	0,0124	0,0150	0,0252	0,0420	0,0580	0,0876
	50	0,0106	0,0116	0,0166	0,0326	0,0582	0,0906	0,1356
	100	0,0142	0,0204	0,0378	0,0810	0,1580	0,2856	0,4162
	200	0,0186	0,0314	0,0832	0,2036	0,4178	0,6592	0,8284
TAD	20	0,0052	0,0072	0,0054	0,0118	0,0128	0,0204	0,0274
	30	0,0090	0,0080	0,0128	0,0162	0,0276	0,0516	0,0670
	40	0,0094	0,0122	0,0150	0,0288	0,0474	0,0706	0,1128
	50	0,0110	0,0114	0,0172	0,0384	0,0706	0,1104	0,1632
	100	0,0140	0,0244	0,0418	0,0970	0,1844	0,3394	0,4848
	200	0,0224	0,0336	0,0950	0,2308	0,4696	0,7120	0,8704

Tabela 59 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0192	0,0168	0,0108	0,0082	0,0094	0,0086	0,0188
	30	0,0112	0,0106	0,0084	0,0130	0,0210	0,0364	0,0524
	40	0,0114	0,0112	0,0152	0,0218	0,0420	0,0718	0,1030
	50	0,0130	0,0112	0,0138	0,0332	0,0694	0,1120	0,1638
	100	0,0134	0,0230	0,0490	0,1150	0,2284	0,3770	0,5246
	200	0,0204	0,0454	0,1472	0,3308	0,5988	0,8042	0,9216
TKS	20	0,0094	0,0084	0,0080	0,0106	0,0158	0,0194	0,0326
	30	0,0080	0,0092	0,0118	0,0184	0,0310	0,0422	0,0566
	40	0,0084	0,0116	0,0162	0,0292	0,0444	0,0716	0,1062
	50	0,0148	0,0150	0,0188	0,0388	0,0702	0,1056	0,1462
	100	0,0162	0,0196	0,0412	0,1000	0,1858	0,3134	0,4274
	200	0,0202	0,0328	0,1004	0,2342	0,4626	0,6932	0,8374
TCM	20	0,0112	0,0100	0,0094	0,0098	0,0118	0,0188	0,0318
	30	0,0076	0,0084	0,0096	0,0174	0,0344	0,0488	0,0734
	40	0,0090	0,0100	0,0164	0,0324	0,0518	0,0890	0,1348
	50	0,0126	0,0168	0,0200	0,0456	0,0862	0,1456	0,2036
	100	0,0166	0,0218	0,0472	0,1232	0,2426	0,4246	0,5726
	200	0,0226	0,0390	0,1318	0,3234	0,6118	0,8394	0,9446
TAD	20	0,0050	0,0060	0,0072	0,0114	0,0192	0,0214	0,0414
	30	0,0060	0,0076	0,0112	0,0234	0,0384	0,0610	0,0942
	40	0,0070	0,0100	0,0192	0,0358	0,0660	0,1120	0,1666
	50	0,0120	0,0154	0,0234	0,0550	0,1004	0,1722	0,2446
	100	0,0170	0,0252	0,0556	0,1462	0,2852	0,4836	0,6444
	200	0,0232	0,0474	0,1510	0,3666	0,6742	0,8790	0,9656

Tabela 60 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 100$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 1% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0128	0,0090	0,0118	0,0114	0,0164	0,0340	0,0438
	30	0,0126	0,0130	0,0198	0,0390	0,0710	0,1050	0,1492
	40	0,0108	0,0160	0,0326	0,0810	0,1426	0,2084	0,2696
	50	0,0130	0,0148	0,0486	0,1342	0,2168	0,3366	0,4192
	100	0,0148	0,0558	0,1850	0,4318	0,6544	0,8128	0,9000
	200	0,0276	0,1308	0,4810	0,8524	0,9700	0,9968	0,9996
TKS	20	0,0108	0,0092	0,0126	0,0182	0,0298	0,0396	0,0566
	30	0,0102	0,0126	0,0230	0,0498	0,0816	0,1146	0,1406
	40	0,0124	0,0128	0,0376	0,0816	0,1466	0,1970	0,2520
	50	0,0108	0,0212	0,0492	0,1260	0,1896	0,2930	0,3536
	100	0,0138	0,0470	0,1532	0,3474	0,5478	0,7160	0,8186
	200	0,0214	0,0932	0,3740	0,7244	0,9266	0,9852	0,9962
TCM	20	0,0074	0,0074	0,0124	0,0198	0,0284	0,0452	0,0674
	30	0,0086	0,0134	0,0236	0,0552	0,1070	0,1416	0,1896
	40	0,0124	0,0166	0,0424	0,1044	0,1846	0,2688	0,3484
	50	0,0138	0,0186	0,0582	0,1642	0,2738	0,4042	0,4916
	100	0,0146	0,0534	0,2070	0,4648	0,7148	0,8646	0,9434
	200	0,0296	0,1208	0,4932	0,8738	0,9868	0,9988	0,9998
TAD	20	0,0052	0,0064	0,0158	0,0268	0,0404	0,0670	0,0916
	30	0,0066	0,0144	0,0308	0,0712	0,1326	0,1894	0,2400
	40	0,0106	0,0212	0,0520	0,1268	0,2242	0,3328	0,4126
	50	0,0140	0,0228	0,0734	0,2002	0,3258	0,4788	0,5910
	100	0,0142	0,0596	0,2332	0,5320	0,7792	0,9156	0,9688
	200	0,0280	0,1380	0,5532	0,9104	0,9930	0,9996	1,0000

Nível nominal de $\alpha = 5\%$

Tabela 63 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0704	0,0738	0,0696	0,0612	0,0640	0,0630	0,0618
	30	0,0574	0,0564	0,0536	0,0588	0,0564	0,0578	0,0576
	40	0,0594	0,0550	0,0580	0,0508	0,0512	0,0502	0,0470
	50	0,0560	0,0550	0,0446	0,0510	0,0588	0,0510	0,0512
	100	0,0528	0,0480	0,0562	0,0574	0,0554	0,0654	0,0612
	200	0,0586	0,0622	0,0728	0,0612	0,0802	0,0822	0,0862
TKS	20	0,0502	0,0502	0,0514	0,0442	0,0486	0,0472	0,0484
	30	0,0516	0,0522	0,0478	0,0550	0,0480	0,0530	0,0528
	40	0,0492	0,0466	0,0524	0,0472	0,0530	0,0504	0,0508
	50	0,0542	0,0500	0,0494	0,0476	0,0540	0,0538	0,0524
	100	0,0540	0,0510	0,0604	0,0620	0,0596	0,0528	0,0660
	200	0,0638	0,0606	0,0662	0,0668	0,0694	0,0714	0,0892
TCM	20	0,0552	0,0526	0,0548	0,0480	0,0478	0,0514	0,0470
	30	0,0550	0,0568	0,0472	0,0520	0,0486	0,0478	0,0528
	40	0,0488	0,0524	0,0536	0,0470	0,0502	0,0546	0,0494
	50	0,0522	0,0478	0,0528	0,0526	0,0516	0,0542	0,0514
	100	0,0590	0,0528	0,0610	0,0594	0,0628	0,0584	0,0674
	200	0,0598	0,0606	0,0666	0,0622	0,0740	0,0782	0,0914
TAD	20	0,0460	0,0384	0,0414	0,0374	0,0370	0,0410	0,0406
	30	0,0450	0,0480	0,0454	0,0444	0,0468	0,0448	0,0508
	40	0,0470	0,0476	0,0528	0,0462	0,0534	0,0514	0,0514
	50	0,0512	0,0516	0,0514	0,0508	0,0528	0,0572	0,0526
	100	0,0634	0,0508	0,0616	0,0620	0,0636	0,0638	0,0710
	200	0,0616	0,0646	0,0732	0,0648	0,0818	0,0868	0,0974

Tabela 64 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0628	0,0706	0,0654	0,0648	0,0598	0,0540	0,0552
	30	0,0528	0,0580	0,0554	0,0534	0,0576	0,0576	0,0556
	40	0,0514	0,0538	0,0598	0,0536	0,0550	0,0504	0,0562
	50	0,0506	0,0512	0,0500	0,0544	0,0538	0,0588	0,0606
	100	0,0534	0,0552	0,0614	0,0632	0,0670	0,0756	0,0710
	200	0,0742	0,0772	0,0732	0,0810	0,0944	0,1206	0,1424
TKS	20	0,0514	0,0468	0,0488	0,0450	0,0488	0,0462	0,0410
	30	0,0516	0,0556	0,0494	0,0454	0,0520	0,0516	0,0538
	40	0,0544	0,0482	0,0504	0,0524	0,0570	0,0544	0,0568
	50	0,0444	0,0522	0,0492	0,0528	0,0526	0,0608	0,0676
	100	0,0540	0,0604	0,0586	0,0582	0,0712	0,0726	0,0740
	200	0,0714	0,0744	0,0640	0,0792	0,0840	0,1022	0,1126
TCM	20	0,0542	0,0530	0,0542	0,0470	0,0470	0,0456	0,0446
	30	0,0480	0,0564	0,0566	0,0478	0,0520	0,0512	0,0552
	40	0,0520	0,0486	0,0490	0,0510	0,0586	0,0540	0,0552
	50	0,0444	0,0526	0,0480	0,0550	0,0568	0,0610	0,0672
	100	0,0540	0,0626	0,0606	0,0562	0,0686	0,0726	0,0802
	200	0,0716	0,0726	0,0690	0,0808	0,0886	0,1108	0,1342
TAD	20	0,0422	0,0404	0,0424	0,0390	0,0364	0,0422	0,0386
	30	0,0404	0,0498	0,0454	0,0444	0,0518	0,0516	0,0518
	40	0,0518	0,0484	0,0502	0,0512	0,0584	0,0556	0,0610
	50	0,0494	0,0512	0,0530	0,0544	0,0534	0,0612	0,0726
	100	0,0542	0,0640	0,0650	0,0610	0,0726	0,0818	0,0864
	200	0,0730	0,0756	0,0746	0,0868	0,0974	0,1294	0,1506

Tabela 65 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0686	0,0652	0,0606	0,0588	0,0596	0,0474	0,0460
	30	0,0602	0,0570	0,0582	0,0514	0,0586	0,0454	0,0602
	40	0,0566	0,0602	0,0516	0,0520	0,0508	0,0694	0,0774
	50	0,0584	0,0544	0,0570	0,0550	0,0676	0,0836	0,1040
	100	0,0508	0,0576	0,0662	0,0892	0,1148	0,1590	0,2234
	200	0,0750	0,0878	0,1116	0,1590	0,2402	0,3500	0,4748
TKS	20	0,0468	0,0446	0,0488	0,0490	0,0498	0,0478	0,0486
	30	0,0504	0,0504	0,0522	0,0506	0,0574	0,0658	0,0682
	40	0,0530	0,0478	0,0536	0,0544	0,0550	0,0710	0,0876
	50	0,0526	0,0524	0,0532	0,0638	0,0712	0,0820	0,1070
	100	0,0566	0,0632	0,0654	0,0900	0,1064	0,1396	0,1994
	200	0,0658	0,0826	0,0998	0,1316	0,2038	0,2704	0,3800
TCM	20	0,0520	0,0482	0,0496	0,0514	0,0472	0,0446	0,0498
	30	0,0490	0,0506	0,0558	0,0544	0,0552	0,0630	0,0692
	40	0,0530	0,0520	0,0518	0,0580	0,0600	0,0736	0,0938
	50	0,0518	0,0526	0,0562	0,0616	0,0726	0,0888	0,1160
	100	0,0572	0,0642	0,0724	0,0944	0,1150	0,1552	0,2266
	200	0,0698	0,0874	0,1116	0,1422	0,2262	0,3250	0,4590
TAD	20	0,0382	0,0380	0,0352	0,0410	0,0432	0,0456	0,0536
	30	0,0450	0,0454	0,0518	0,0524	0,0636	0,0636	0,0834
	40	0,0480	0,0488	0,0490	0,0572	0,0650	0,0836	0,1082
	50	0,0526	0,0530	0,0588	0,0650	0,0792	0,1024	0,1324
	100	0,0562	0,0658	0,0776	0,1006	0,1276	0,1808	0,2544
	200	0,0780	0,0920	0,1180	0,1604	0,2456	0,3640	0,5020

Tabela 66 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0682	0,0688	0,0664	0,0486	0,0482	0,0434	0,0450
	30	0,0630	0,0552	0,0530	0,0468	0,0540	0,0616	0,0896
	40	0,0554	0,0604	0,0514	0,0654	0,0718	0,0972	0,1324
	50	0,0486	0,0526	0,0600	0,0704	0,0968	0,1410	0,1818
	100	0,0664	0,0670	0,0878	0,1382	0,2108	0,3338	0,4370
	200	0,0790	0,0944	0,1654	0,2720	0,4442	0,6344	0,7860
TKS	20	0,0540	0,0500	0,0472	0,0418	0,0546	0,0536	0,0576
	30	0,0472	0,0504	0,0538	0,0532	0,0638	0,0740	0,1014
	40	0,0504	0,0540	0,0556	0,0742	0,0846	0,1058	0,1376
	50	0,0550	0,0540	0,0612	0,0828	0,1036	0,1372	0,1786
	100	0,0668	0,0736	0,0848	0,1208	0,1912	0,2782	0,3742
	200	0,0714	0,0882	0,1352	0,2278	0,3716	0,5328	0,6912
TCM	20	0,0540	0,0502	0,0472	0,0446	0,0494	0,0542	0,0548
	30	0,0496	0,0520	0,0576	0,0568	0,0634	0,0754	0,1036
	40	0,0496	0,0556	0,0570	0,0700	0,0890	0,1194	0,1544
	50	0,0536	0,0558	0,0640	0,0752	0,1066	0,1562	0,1982
	100	0,0692	0,0710	0,0834	0,1398	0,2134	0,3284	0,4440
	200	0,0748	0,0940	0,1498	0,2636	0,4338	0,6296	0,7904
TAD	20	0,0374	0,0406	0,0424	0,0382	0,0518	0,0564	0,0650
	30	0,0452	0,0446	0,0516	0,0578	0,0736	0,0874	0,1274
	40	0,0466	0,0500	0,0578	0,0796	0,1004	0,1418	0,1830
	50	0,0510	0,0580	0,0630	0,0828	0,1218	0,1796	0,2358
	100	0,0696	0,0748	0,0892	0,1574	0,2442	0,3742	0,4978
	200	0,0794	0,1012	0,1650	0,2902	0,4774	0,6816	0,8336

Tabela 67 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0738	0,0638	0,0508	0,0474	0,0474	0,0522	0,0568
	30	0,0612	0,0504	0,0522	0,0534	0,0686	0,0968	0,1200
	40	0,0548	0,0568	0,0606	0,0714	0,1040	0,1520	0,1934
	50	0,0554	0,0540	0,0596	0,1026	0,1430	0,2010	0,2670
	100	0,0600	0,0718	0,1156	0,2086	0,3382	0,4896	0,6310
	200	0,0802	0,1136	0,2258	0,4246	0,6586	0,8342	0,9260
TKS	20	0,0502	0,0486	0,0456	0,0542	0,0520	0,0680	0,0792
	30	0,0512	0,0464	0,0566	0,0604	0,0752	0,1098	0,1502
	40	0,0498	0,0552	0,0628	0,0826	0,1152	0,1496	0,2026
	50	0,0572	0,0586	0,0678	0,0970	0,1450	0,1978	0,2684
	100	0,0616	0,0754	0,1074	0,1816	0,2778	0,4272	0,5554
	200	0,0704	0,1012	0,1842	0,3442	0,5528	0,7434	0,8754
TCM	20	0,0520	0,0494	0,0438	0,0512	0,0514	0,0660	0,0814
	30	0,0536	0,0452	0,0580	0,0612	0,0806	0,1192	0,1614
	40	0,0522	0,0556	0,0646	0,0858	0,1236	0,1672	0,2308
	50	0,0504	0,0624	0,0688	0,1066	0,1650	0,2300	0,3104
	100	0,0606	0,0786	0,1182	0,2050	0,3386	0,5138	0,6592
	200	0,0782	0,1124	0,2054	0,4122	0,6510	0,8430	0,9440
TAD	20	0,0386	0,0412	0,0378	0,0498	0,0570	0,0814	0,0972
	30	0,0466	0,0438	0,0594	0,0692	0,0954	0,1444	0,1910
	40	0,0484	0,0574	0,0744	0,0944	0,1454	0,1994	0,2720
	50	0,0492	0,0612	0,0748	0,1244	0,1912	0,2660	0,3584
	100	0,0604	0,0850	0,1294	0,2350	0,3846	0,5614	0,7092
	200	0,0816	0,1172	0,2276	0,4520	0,7012	0,8812	0,9618

Tabela 68 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, 5% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,0718	0,0632	0,0490	0,0428	0,0498	0,0496	0,0724
	30	0,0594	0,0536	0,0464	0,0634	0,0768	0,1246	0,1716
	40	0,0550	0,0514	0,0628	0,0874	0,1356	0,1948	0,2770
	50	0,0538	0,0470	0,0688	0,1146	0,2026	0,2832	0,3660
	100	0,0524	0,0844	0,1532	0,2848	0,4370	0,6232	0,7656
	200	0,0778	0,1292	0,3168	0,5570	0,8124	0,9340	0,9818
TKS	20	0,0476	0,0474	0,0438	0,0522	0,0588	0,0752	0,0952
	30	0,0508	0,0520	0,0538	0,0748	0,0976	0,1356	0,1780
	40	0,0540	0,0552	0,0684	0,0936	0,1416	0,1954	0,2694
	50	0,0564	0,0606	0,0744	0,1230	0,1872	0,2634	0,3494
	100	0,0636	0,0744	0,1332	0,2438	0,3844	0,5538	0,6784
	200	0,0790	0,1206	0,2438	0,4626	0,7088	0,8770	0,9520
TCM	20	0,0500	0,0504	0,0448	0,0520	0,0624	0,0758	0,1038
	30	0,0506	0,0468	0,0570	0,0782	0,1076	0,1462	0,2058
	40	0,0518	0,0524	0,0706	0,0998	0,1606	0,2258	0,3108
	50	0,0576	0,0598	0,0774	0,1372	0,2138	0,3166	0,4176
	100	0,0628	0,0760	0,1450	0,2842	0,4622	0,6528	0,7884
	200	0,0778	0,1254	0,2902	0,5538	0,8100	0,9458	0,9860
TAD	20	0,0432	0,0424	0,0444	0,0520	0,0694	0,0928	0,1244
	30	0,0424	0,0472	0,0592	0,0894	0,1248	0,1840	0,2542
	40	0,0476	0,0492	0,0756	0,1186	0,1822	0,2630	0,3720
	50	0,0558	0,0568	0,0842	0,1552	0,2518	0,3662	0,4780
	100	0,0636	0,0848	0,1690	0,3244	0,5116	0,7124	0,8368
	200	0,0830	0,1340	0,3256	0,6018	0,8566	0,9680	0,9928

Nível nominal de $\alpha = 10\%$

Tabela 71 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 8$, $n = 20, 30, 40, 50, 100, 200$, nível nominal de 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1294	0,1356	0,1306	0,1164	0,1166	0,1130	0,1194
	30	0,1082	0,1122	0,1052	0,1106	0,1046	0,1128	0,1042
	40	0,1064	0,1024	0,1122	0,1028	0,1040	0,1044	0,0986
	50	0,1074	0,1012	0,1020	0,1028	0,1086	0,0978	0,1014
	100	0,1066	0,1000	0,1092	0,1090	0,1118	0,1160	0,1220
	200	0,1182	0,1210	0,1260	0,1216	0,1398	0,1422	0,1556
TKS	20	0,0958	0,1002	0,1046	0,0916	0,0916	0,0962	0,0998
	30	0,1034	0,1074	0,0990	0,1014	0,0996	0,0964	0,1026
	40	0,0998	0,0974	0,1052	0,0946	0,1074	0,0998	0,0972
	50	0,1014	0,1054	0,1008	0,1038	0,1072	0,1058	0,1022
	100	0,1122	0,1028	0,1098	0,1120	0,1126	0,1092	0,1220
	200	0,1194	0,1140	0,1216	0,1220	0,1300	0,1378	0,1554
TCM	20	0,0966	0,1042	0,1040	0,0950	0,0982	0,1012	0,1008
	30	0,1030	0,1090	0,1014	0,1068	0,1022	0,0938	0,0994
	40	0,1000	0,1012	0,1116	0,0984	0,1052	0,1032	0,0994
	50	0,0998	0,0990	0,1028	0,1026	0,1048	0,1100	0,1042
	100	0,1132	0,0980	0,1120	0,1220	0,1166	0,1176	0,1256
	200	0,1146	0,1220	0,1252	0,1248	0,1340	0,1440	0,1568
TAD	20	0,0864	0,0888	0,0904	0,0864	0,0848	0,0918	0,0860
	30	0,0972	0,0970	0,0892	0,0980	0,0936	0,0946	0,0932
	40	0,0944	0,0982	0,1066	0,0938	0,1090	0,0988	0,1008
	50	0,1004	0,0964	0,1022	0,0992	0,1098	0,1124	0,1058
	100	0,1172	0,1024	0,1210	0,1216	0,1184	0,1252	0,1274
	200	0,1208	0,1220	0,1304	0,1274	0,1434	0,1520	0,1690

Tabela 72 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 10$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1288	0,1272	0,1240	0,1218	0,1106	0,1094	0,1104
	30	0,1004	0,1164	0,1096	0,1030	0,1078	0,1090	0,1092
	40	0,1072	0,1046	0,1062	0,1088	0,1018	0,1030	0,1028
	50	0,0988	0,1010	0,0998	0,1024	0,1112	0,1132	0,1128
	100	0,1048	0,1108	0,1148	0,1226	0,1262	0,1358	0,1364
	200	0,1322	0,1362	0,1334	0,1416	0,1630	0,2040	0,2276
TKS	20	0,0998	0,1012	0,1010	0,0918	0,0954	0,0916	0,0926
	30	0,1042	0,1078	0,0996	0,0944	0,1040	0,0982	0,1028
	40	0,1016	0,0988	0,1036	0,1058	0,1106	0,1040	0,1108
	50	0,0922	0,1058	0,1004	0,1088	0,1048	0,1210	0,1220
	100	0,1096	0,1114	0,1128	0,1094	0,1272	0,1358	0,1370
	200	0,1280	0,1328	0,1194	0,1414	0,1548	0,1758	0,1972
TCM	20	0,0982	0,1020	0,1046	0,0952	0,0924	0,0926	0,0938
	30	0,1032	0,1090	0,1026	0,0922	0,1042	0,1014	0,1030
	40	0,1020	0,0986	0,1010	0,1046	0,1112	0,1050	0,1114
	50	0,0948	0,1058	0,1016	0,1058	0,1096	0,1218	0,1220
	100	0,1114	0,1150	0,1176	0,1120	0,1270	0,1354	0,1410
	200	0,1316	0,1334	0,1246	0,1496	0,1554	0,1920	0,2140
TAD	20	0,0838	0,0880	0,0858	0,0864	0,0836	0,0844	0,0850
	30	0,0932	0,1016	0,0970	0,0904	0,1024	0,1024	0,1088
	40	0,0986	0,0992	0,1022	0,1056	0,1074	0,1066	0,1176
	50	0,0906	0,1008	0,1020	0,1040	0,1084	0,1252	0,1240
	100	0,1146	0,1180	0,1188	0,1214	0,1378	0,1442	0,1546
	200	0,1398	0,1400	0,1340	0,1574	0,1678	0,2030	0,2370

Tabela 73 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 20$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1310	0,1256	0,1188	0,1172	0,1072	0,0974	0,0916
	30	0,1126	0,1078	0,1110	0,1040	0,1078	0,1006	0,1148
	40	0,1098	0,1120	0,0990	0,1058	0,1002	0,1254	0,1364
	50	0,1110	0,1060	0,1088	0,1078	0,1200	0,1478	0,1844
	100	0,0976	0,1158	0,1248	0,1510	0,1938	0,2476	0,3294
	200	0,1352	0,1472	0,1858	0,2430	0,3434	0,4680	0,5968
TKS	20	0,1002	0,0972	0,0982	0,0960	0,1004	0,0968	0,1048
	30	0,0966	0,1012	0,0996	0,1034	0,1126	0,1182	0,1258
	40	0,1006	0,0990	0,1088	0,1104	0,1142	0,1328	0,1488
	50	0,1076	0,1066	0,1060	0,1188	0,1316	0,1516	0,1832
	100	0,1108	0,1140	0,1294	0,1582	0,1836	0,2320	0,3012
	200	0,1244	0,1442	0,1732	0,2128	0,2992	0,3876	0,5140
TCM	20	0,1002	0,0962	0,1022	0,1016	0,0960	0,0962	0,0986
	30	0,1008	0,1012	0,1052	0,1048	0,1126	0,1184	0,1278
	40	0,1024	0,1002	0,1022	0,1118	0,1164	0,1338	0,1594
	50	0,1096	0,1032	0,1142	0,1182	0,1316	0,1462	0,1926
	100	0,1124	0,1142	0,1316	0,1590	0,1948	0,2496	0,3270
	200	0,1292	0,1486	0,1894	0,2256	0,3280	0,4450	0,5812
TAD	20	0,0864	0,0848	0,0850	0,0882	0,0912	0,0918	0,1028
	30	0,0930	0,0930	0,1002	0,1076	0,1146	0,1210	0,1444
	40	0,0968	0,1014	0,0984	0,1152	0,1196	0,1504	0,1738
	50	0,1030	0,0988	0,1132	0,1198	0,1424	0,1736	0,2186
	100	0,1076	0,1252	0,1352	0,1734	0,2156	0,2730	0,3628
	200	0,1356	0,1542	0,1990	0,2510	0,3524	0,4824	0,6254

Tabela 74 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 30$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1252	0,1286	0,1238	0,1022	0,0974	0,0880	0,0948
	30	0,1160	0,1186	0,1046	0,0978	0,1052	0,1158	0,1558
	40	0,1082	0,1140	0,1072	0,1180	0,1348	0,1676	0,2236
	50	0,1052	0,1034	0,1106	0,1300	0,1690	0,2298	0,2806
	100	0,1224	0,1214	0,1548	0,2108	0,3160	0,4482	0,5662
	200	0,1372	0,1592	0,2514	0,3894	0,5682	0,7438	0,8640
TKS	20	0,1042	0,0980	0,0960	0,0876	0,1026	0,1008	0,1100
	30	0,1050	0,0986	0,1002	0,1054	0,1208	0,1362	0,1728
	40	0,0998	0,1068	0,1136	0,1314	0,1476	0,1820	0,2254
	50	0,1058	0,1060	0,1208	0,1374	0,1700	0,2202	0,2776
	100	0,1216	0,1290	0,1428	0,2052	0,2924	0,3954	0,4930
	200	0,1350	0,1536	0,2172	0,3236	0,5026	0,6516	0,7996
TCM	20	0,1104	0,1024	0,0964	0,0888	0,0988	0,1016	0,1124
	30	0,1004	0,1054	0,1042	0,1100	0,1162	0,1380	0,1760
	40	0,0982	0,1110	0,1112	0,1322	0,1482	0,1940	0,2414
	50	0,1044	0,1070	0,1210	0,1400	0,1822	0,2402	0,2980
	100	0,1204	0,1286	0,1534	0,2188	0,3194	0,4450	0,5598
	200	0,1336	0,1572	0,2420	0,3680	0,5496	0,7354	0,8682
TAD	20	0,0892	0,0878	0,0896	0,0800	0,1002	0,1070	0,1254
	30	0,0922	0,0970	0,1032	0,1126	0,1294	0,1552	0,2058
	40	0,0990	0,1078	0,1110	0,1440	0,1644	0,2238	0,2762
	50	0,1056	0,1074	0,1266	0,1460	0,2052	0,2722	0,3448
	100	0,1276	0,1370	0,1590	0,2400	0,3530	0,4908	0,6194
	200	0,1364	0,1692	0,2618	0,3996	0,5994	0,7804	0,8982

Tabela 75 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 40$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1326	0,1218	0,0994	0,1024	0,0964	0,1014	0,1104
	30	0,1156	0,1052	0,1016	0,1106	0,1230	0,1654	0,2058
	40	0,1060	0,1054	0,1128	0,1296	0,1784	0,2330	0,2994
	50	0,1022	0,1052	0,1150	0,1734	0,2256	0,3102	0,3962
	100	0,1118	0,1320	0,1894	0,3094	0,4502	0,6104	0,7488
	200	0,1474	0,1892	0,3328	0,5374	0,7618	0,9000	0,9618
TKS	20	0,0998	0,0964	0,0964	0,1038	0,1032	0,1230	0,1374
	30	0,1006	0,0936	0,1120	0,1142	0,1376	0,1872	0,2310
	40	0,1006	0,1074	0,1182	0,1444	0,1824	0,2344	0,3018
	50	0,1086	0,1126	0,1222	0,1704	0,2298	0,3012	0,3750
	100	0,1154	0,1364	0,1824	0,2708	0,4038	0,5652	0,6884
	200	0,1334	0,1744	0,2826	0,4642	0,6840	0,8446	0,9282
TCM	20	0,1028	0,1022	0,0890	0,1014	0,1012	0,1230	0,1406
	30	0,1044	0,0968	0,1168	0,1148	0,1454	0,1942	0,2526
	40	0,0996	0,1050	0,1246	0,1456	0,1924	0,2542	0,3354
	50	0,1042	0,1120	0,1260	0,1822	0,2494	0,3274	0,4236
	100	0,1144	0,1382	0,1874	0,3006	0,4528	0,6276	0,7582
	200	0,1416	0,1842	0,3080	0,5324	0,7630	0,9052	0,9732
TAD	20	0,0848	0,0874	0,0830	0,1016	0,1074	0,1386	0,1632
	30	0,0970	0,0910	0,1146	0,1256	0,1626	0,2266	0,2942
	40	0,0958	0,1082	0,1334	0,1626	0,2214	0,2970	0,3822
	50	0,1010	0,1158	0,1350	0,2066	0,2812	0,3784	0,4802
	100	0,1154	0,1438	0,2084	0,3406	0,5000	0,6772	0,8038
	200	0,1452	0,2010	0,3362	0,5742	0,7964	0,9336	0,9826

Tabela 76 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 50$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1310	0,1182	0,0988	0,0902	0,1052	0,1014	0,1340
	30	0,1192	0,1090	0,0996	0,1184	0,1440	0,2084	0,2716
	40	0,1070	0,1092	0,1210	0,1580	0,2222	0,2922	0,4026
	50	0,1046	0,1002	0,1286	0,1938	0,2962	0,4038	0,4996
	100	0,1096	0,1478	0,2430	0,3954	0,5650	0,7380	0,8568
	200	0,1412	0,2140	0,4248	0,6820	0,8844	0,9676	0,9908
TKS	20	0,1010	0,0966	0,0962	0,0996	0,1146	0,1408	0,1638
	30	0,0996	0,1040	0,1104	0,1318	0,1612	0,2202	0,2834
	40	0,1046	0,1030	0,1304	0,1592	0,2228	0,2958	0,3890
	50	0,1088	0,1144	0,1366	0,2018	0,2938	0,3848	0,4818
	100	0,1200	0,1372	0,2214	0,3454	0,5178	0,6762	0,7952
	200	0,1276	0,1964	0,3522	0,5954	0,8170	0,9344	0,9794
TCM	20	0,1034	0,1040	0,0964	0,1020	0,1178	0,1412	0,1740
	30	0,0998	0,0984	0,1090	0,1398	0,1774	0,2376	0,3032
	40	0,1004	0,1036	0,1348	0,1732	0,2422	0,3258	0,4340
	50	0,1074	0,1124	0,1410	0,2130	0,3134	0,4372	0,5434
	100	0,1198	0,1396	0,2362	0,3898	0,5852	0,7608	0,8708
	200	0,1400	0,2036	0,3992	0,6670	0,8842	0,9730	0,9944
TAD	20	0,0870	0,0896	0,0876	0,0982	0,1328	0,1608	0,2032
	30	0,0954	0,0978	0,1160	0,1544	0,2008	0,2852	0,3634
	40	0,0950	0,1052	0,1398	0,1962	0,2810	0,3766	0,5000
	50	0,1034	0,1092	0,1500	0,2394	0,3592	0,4898	0,6002
	100	0,1166	0,1526	0,2620	0,4348	0,6342	0,8060	0,9050
	200	0,1474	0,2178	0,4382	0,7094	0,9116	0,9830	0,9978

Tabela 77 – Taxas de rejeição de H_0 : os dados possuem distribuição Bingham complexa, quando H_0 é falsa. Utilizando os testes TRV, TKS, TCM e TAD pelo método *bootstrap*, para $\sigma = 100$, $n = 20, 30, 40, 50, 100, 200$, 10% e 7 valores de contaminação.

Teste	n	ϵ						
		ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7
TRV	20	0,1232	0,1050	0,1000	0,1118	0,1446	0,1992	0,2488
	30	0,1142	0,1086	0,1372	0,2234	0,3154	0,3936	0,4868
	40	0,1058	0,1178	0,1896	0,3240	0,4740	0,5650	0,6596
	50	0,1132	0,1338	0,2516	0,4410	0,5882	0,7166	0,8026
	100	0,1124	0,2338	0,5156	0,7848	0,9188	0,9732	0,9890
	200	0,1660	0,4094	0,8130	0,9728	0,9980	0,9998	1,0000
TKS	20	0,0962	0,0986	0,1128	0,1402	0,1860	0,2284	0,2736
	30	0,0932	0,1146	0,1556	0,2346	0,3294	0,4054	0,4766
	40	0,0992	0,1160	0,2004	0,3216	0,4684	0,5568	0,6316
	50	0,1152	0,1338	0,2384	0,4164	0,5584	0,6740	0,7474
	100	0,1108	0,2236	0,4596	0,7134	0,8768	0,9502	0,9798
	200	0,1620	0,3566	0,7362	0,9532	0,9956	0,9988	1,0000
TCM	20	0,1050	0,0978	0,1080	0,1408	0,1908	0,2374	0,2968
	30	0,1014	0,1198	0,1566	0,2488	0,3686	0,4524	0,5440
	40	0,1048	0,1174	0,2148	0,3494	0,5114	0,6276	0,7006
	50	0,1130	0,1366	0,2646	0,4684	0,6370	0,7664	0,8362
	100	0,1098	0,2382	0,5104	0,7946	0,9312	0,9840	0,9964
	200	0,1678	0,4002	0,8102	0,9808	0,9990	1,0000	1,0000
TAD	20	0,0854	0,0902	0,1190	0,1678	0,2350	0,2950	0,3638
	30	0,0922	0,1258	0,1782	0,2954	0,4252	0,5212	0,6254
	40	0,1010	0,1350	0,2432	0,4038	0,5782	0,6972	0,7834
	50	0,1164	0,1506	0,3022	0,5308	0,7024	0,8246	0,8910
	100	0,1178	0,2660	0,5632	0,8418	0,9546	0,9918	0,9984
	200	0,1844	0,4322	0,8514	0,9900	0,9996	1,0000	1,0000

