Universidade Federal de Pernambuco

Centro de Ciências Exatas e da Natureza

Departamento de Estatística

Programa de Pós-Graduação em Estatística

Wenia Valdevino Félix

Statistical Inference Based on Information Theory for Pre-shape Data

Recife

2019

Wenia Valdevino Félix

Statistical Inference Based on Information Theory for Pre-shape Data

Doctoral thesis submitted to the Programa de Pós-Graduação em Estatística na Universidade Federal de Pernambuco as a partial requirement for obtaining a doctorate in Statistics.

Concentration area: Mathematic Statistics.

Advisor: Prof. Dr. Abraão D. C. Nascimento
Co-advisor: Prof. Dr. Getúlio J. A. do Amaral

Recife
2019

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

F316s Félix, Wenia Valdevino
Statistical inference based on information theory for pre-shape data / Wenia Valdevino Félix. – 2019.
136 f.: il., fig., tab.

Orientador: Abraão D. C. Nascimento.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2019.
Inclui referências e apêndices.

1. Estatística. 2. Distâncias estocásticas. I. Nascimento, Abraão D. C. (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2019-032

WENIA VALDEVINO FÉLIX

STATISTICAL INFERENCE BASED ON INFORMATION THEORY FOR PRE-SHAPE DATA

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

Aprovada em: 27 de fevereiro de 2019.

BANCA EXAMINADORA

Prof.(º) Abraão David Costa do Nascimento
UFPE

Prof.(ª) Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof.(º) Alex Dias Ramos
UFPE

Prof.(º) Eufrásio de Andrade Lima Neto
UFPB

Prof.(º) Paulo José Duarte Neto
UFRPE

Dedico este trabalho à Deus,

a meus pais, João Félix e Elizabeth Valdevino

e a meu esposo, Romildo Lima.

AGRADECIMENTOS

Agradeço a Deus pela oportunidade de ingressar neste doutorado e por me auxiliar nos momentos difíceis, que com sua graça, foram superados.

Aos meus orientadores, Dr. Abraão David e Dr. Getúlio Amorim, por toda a paciência, incentivo, compreensão, disponibilidade, confiança e de forma geral por todos os conhecimentos científicos e lições de vida que me proporcionaram durante esses quatro anos. Além de serem profissionais muito competentes, sempre transmitindo com muita seriedade e naturalidade o conhecimento científico. Em especial, por me apresentar as duas linhas de pesquisa, a saber, Análise Estatística de Formas e Teoria Estatística da Informação. Tópicos extremamente interessantes, tanto no aspecto teórico quanto estatístico e computacional.

A minha família, minha mãe Elizabeth Ribeiro, que por quatro anos, apesar da distância, esteve sempre ao meu lado me apoiando. A meu pai, João Félix, que estava sempre em oração e contando os dias para que pudesse conquistar mais essa vitória. A minha tia Diomar Ribeiro, por todo incentivo e encorajamento que me proporcionou. A meu esposo, Romildo Lima, pela paciência, carinho, compreensão e auxílio durante este longo tempo de dedicação acadêmica.

Aos irmãos e amigos da igreja Missão Evangélica a qual fazia parte em Natal, em especial ao casal Márcia Tavares e Paulo Morais que me auxiliaram diretamente nesses dez anos de carreira acadêmica. Como também a família do Pastor Gonzaga, pelas orações e incentivo.

Aos meus amigos e companheiros do doutorado e mestrado por todos os momentos de estudo e descontração que tivemos. Em especial agradeço aos colegas da minha turma de doutorado, Luana Cecília, Thiago Vedovatto e Marley Apolinário (sempre marlerizando). Estendo meus agradecimentos também ao meu amigo Josimar Vasconcelos, grande estatístico, o qual até hoje me socorre quando o procuro. Como também a nosso representante de doutorado César Diogo e a sua parceira inseparável Maria Cristina. Lembro-me das madrugadas de estudo, das pescadas sobre o notebook e também dos momentos de diversão que tivemos.

Ao professor Dr. Alex Dias por ter aceitado me supervisionar os dois semestres de estágio à docência que foram realizados na disciplina de Introdução à Análise e Probabilidade IV. Período de muito aprendizado para minha carreira docente.

A todos os professores do departamento de Estatística da UFPE, pelos conhecimentos que me foram passados. Aos professores que fizeram parte da minha banca examinadora de qualificação, a saber, Dra. Audrey Cysneiros e Dr. Alex Dias. Aos participantes da banca examinadora da defesa, agradeço pelos comentários e sugestões.

A Valéria Bittencourt, secretária do programa de pós-graduação em Estatística, pela imensa competência e o cuidado no momento de me auxiliar em burocracias acadêmicas. Como também a todos os funcionários do departamento que direta e indiretamente nos auxiliaram no decorrer do curso.

A agência de fomento que financiou o desenvolvimento deste trabalho, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

ABSTRACT

An important branch at Multivariate Analysis is Statistical Shape Analysis (SSA). A common demand in SSA is to study the shape property over objects in images, called planar-shape. To quantify difference in planar-shape between distinct groups is crucial in several areas – such as biology, medical image analysis, among others – and we provide advances in this sense. This thesis assumes pre-shapes which are obtained from two-dimensional objects follow the complex Bingham ($\mathbb{C}B$) distribution, having like the most important particular case the complex Watson ($\mathbb{C}W$) model. From numerical evidence we present in this thesis, statistical tests which are well-defined in the SSA literature may provide low empirical test power curves. In order to obtain new alternatives to overcome this issue, we use information theory measures; in particular, stochastic entropies and distances. These measures play an important role in statistical theory, specifically into estimation and hypothesis inference procedures under large samples. First, we propose new distance-based two-sample hypothesis tests for triangle mean shapes. Closed-form expressions for the Rényi, Kullback-Leibler (KL), Bhattacharyya and Hellinger distances for the $\mathbb{C}W$ distribution are derived. The performance of proposed tests is quantified and compared with that due to the F2 test (analysis-of-variance tailored to the SSA literature). Furthermore, we perform an application to real data. Second, we extend the first topic proposing new distance-based two-sample hypothesis tests (for both homogeneity and mean shape) for the $\mathbb{C}B$ distribution and landmarks number higher than three. We derive the Rényi and KL divergences and the Bhattacharyya and Hellinger distances for the $\mathbb{C}B$ distribution. Three from among them may also be used like tests between two mean shapes or as discrepancy measures between the $\mathbb{C}B$ models. We prove also that the KL discrepancy for the $\mathbb{C}B$ model is rotation invariant. In order to evaluate and compare our proposals with other four SSA mean shape tests, a simulation study is also made to evaluate asymptotic and robustness properties. Finally an application in evolutionary biology is made. Third we tackle the proposal of new entropy-based multi-sample tests for variability in planar-shape. We develop closed-form expressions for the Rényi and Shannon entropies at the $\mathbb{C}B$ and $\mathbb{C}W$ models. From these quantities, hypothesis tests

are obtained to assess if multiple spherical samples have the same degree of disorder.

Keywords: Stochastics distance. Complex Watson distribution. Complex Bingham model. Shannon entropy. Rényi entropy. Mean shape.

RESUMO

Um ramo importante na Análise Multivariada é a Análise Estatística da Forma (AEF). Uma demanda comum em AEF é estudar propriedades de forma sobre objetos em imagens, chamada forma-planar. Quantificar diferenças em forma-planar entre grupos distintos é crucial em várias áreas – tais como biologia, análise de imagens médicas e outras – e nós apresentamos avanços neste sentido. Esta tese pressupõe que pré-formas obtidas a partir de objetos bidimensionais seguem a distribuição Bingham complexa ($\mathbb{C}B$), tendo como o mais importante caso particular o modelo Watson complexo ($\mathbb{C}W$). A partir de evidências numéricas que apresentamos nesta tese, testes bem definidos na literatura da AEF podem fornecer baixas curvas empíricas para o poder do teste. A fim de obter novas alternativas para superar este problema, usamos medidas da Teoria da Informação; em particular, entropias e distâncias estocásticas. Essas medidas desempenham um papel importante na teoria estatística, especificamente nos procedimentos de inferência por estimação e hipóteses em grandes amostras. Primeiro, propomos novos testes de hipóteses de duas amostras baseados em distâncias para formas médias de triângulo. Expressões em forma fechada para as distâncias de Rényi, Kullback-Leibler (KL), Bhattacharyya e Hellinger para a distribuição $\mathbb{C}W$ são derivadas. O desempenho dos testes propostos é quantificado e comparado com o teste F2 (análise de variância adaptada à literatura de AEF). Além disso, realizamos uma aplicação a dados reais. Em segundo lugar, estendemos o primeiro tópico propondo novos testes de hipóteses de duas amostras baseados em distância (tanto para homogeneidade quanto para forma média) para a distribuição $\mathbb{C}B$ e número de pontos de referência maior que três. Derivamos as divergências de Rényi e KL e as distâncias de Bhattacharyya e Hellinger para a distribuição $\mathbb{C}B$. Três dentre elas também podem ser usadas como testes entre duas formas médias ou como medidas de discrepância entre modelos $\mathbb{C}B$. Provamos também que a divergência KL para o modelo $\mathbb{C}B$ é invariante à rotação. Para avaliar e comparar as nossas propostas com outros quatro testes de forma média em AEF, também é feito um estudo de simulação para avaliar propriedades assintóticas e de robustez. Finalmente, uma aplicação em biologia da evolução é feita. Em terceiro lugar, abordamos a proposta de novos testes de múltiplas amostras baseados

em entropia para a variabilidade em forma-planar. Desenvolvemos expressões em forma fechada para as entropias de Rényi e Shannon nos modelos $\mathbb{C}B$ e $\mathbb{C}W$. A partir dessas quantidades, testes de hipóteses são obtidos para avaliar se múltiplas amostras esféricas têm o mesmo grau de desordem.

Palavras-chave: Distâncias estocásticas. Distribuição Watson complexa. Modelo Bingham complexo, Entropia de Shannon. Entropia de Rényi. Forma média.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

SSA	Statistical Shape Analysis
SIT	Statistical Information Theory
KL	Kullback-Leibler
B	Bhattacharyya
R	Rényi
H	Hellinger
KS	Kollmogorov-Smirnov
FIM	Fisher information matrix
ML	Maximum likelihood
pdf	Probability density function
TME	Truncated multivariate exponential
TCL	Lebesgue dominated convergence theorem
SAR	Synthetic Aperture Radar

SYMBOLS LIST

$\mathbb{R}_+$	Positive real numbers
$\boldsymbol{z}^0$	Original complex landmark
$\boldsymbol{z}_H$	Helmerized complex landmark
$\boldsymbol{z}_C^0$	Centred complex landmark
$\boldsymbol{S}$	Product matrix
κ	Concentration parameter of the complex Watson distribution
$\boldsymbol{\mu}$	Modal vector
k	Landmarks number
$D_{KL}(\cdot,\cdot)$	Kullback-Leibler divergence
$D_R^\beta(\cdot,\cdot)$	Rényi divergence
β	Order parameter of Rényi divergence, distance and entropy
β_{otm}	Optimized order parameter of Rényi entropy
$\mathbb{C}B$	Complex Bingham distribution
$\mathbb{C}W$	Complex Watson distribution
$\boldsymbol{H}$	Helmert matrix
$\boldsymbol{R}$	Rotation matrix
$S_{KL}(\cdot,\cdot)$	Kullback-Leibler statistic
$S_H(\cdot,\cdot)$	Hellinger statistic
$S_B(\cdot,\cdot)$	Bhattacharyya statistic
$S_R(\cdot,\cdot)$	Rényi statistic

$\lambda_{\min}$	Lambda statistic
F_H	Hotelling's T^2 statistic
F_G	Goodall statistic
F_J	James statistic
ρ	Procrustes (or Riemannian) distance
d_F	Full Procrustes distance
$c(\cdot)$	Normalization constant of the complex Bingham distribution
$c_1(\cdot)$	Integration constant of the complex Watson distribution
${}_1F_1(\cdot,\cdot,\cdot)$	Confluent hypergeometric function
$H_S(\cdot)$	Shannon entropy
$H_R^{\beta}(\cdot)$	Rényi entropy
χ^2_m	Chi-squared distribution with m degrees of freedom
$\mathbb{C}N_k(\cdot,\cdot)$	Multivariate complex normal distribution
F_{d_1,d_2}	Snedecor's F-distribution with d_1 and d_2 degrees of freedom
$\mathbb{E}(\cdot)$	Expected value

CONTENTS

1 INTRODUCTION

In this chapter we first motivate this manuscript. In particular, we discuss about a possible combination lines between: Statistical Shape Analysis (SSA) and Statistical Information Theory (SIT). Further main contributions, computational support and structure of this thesis are also furnished.

1.1 Motivation

Over the years, SSA has gained attention in the literature due several areas, such as:

- In Biology: Cranial differences between the sexes of apes – in particular male and female adult gorilla skulls – have been studied by O'Higgins (1989) and O'Higgins and Dryden (1993). Sonat et al. (2009) have proposed landmarks for the hippocampus in rat brains based on empirical experiment and have investigated hippocampus shape changes in rat brain with epilepsy using SSA methods.

- In Medicine: Studies on focal neuroanatomical anomalies in patients with schizophrenia have been made by Bookstein (1996), assessing the shape differences in the brain. In the medical image analysis context, Golland et al. (2005) have presented a computational framework for image-based analysis and interpretation of statistical differences in anatomical shape between populations. Colak et al. (2011) have investigated shape differences of the corpus callosum in patients with Behçet's disease using SSA.

- In Geology: Mean shapes of fossils and structures for their variability have been estimated and studied by Lohmann (1983) and Dryden and Mardia (2016).

- In image detection and processing: Frangi et al. (2002) have provided a generic framework for automatic selection of corresponding landmarks in three-dimension (3D) shapes. Subsequently, this technique has been applied at the construction of 3D shape models for 3D cardiac magnetic resonance imagery. Wang et al. (2003) have furnished a simple and efficient automatic gait recognition algorithm utilizing SSA.

The SSA can be understood as a set of procedures for analyzing shapes under uncertainty (DRYDEN; MARDIA, 2016). The core concept of this theory is the one of *shape*, geometrical information which remains after removing the effects of changes in location, scale and rotation. One manner to work with shapes of objects is to define a geometrical configuration called *landmark configuration*, which labels points (known as landmarks from (i) anatomical, (ii) mathematical and (iii) imputation points of view, (BROMBIN et al., 2016) over them, according to some common coordinate system; for instance, Bookstein (BOOKSTEIN, 1986) and Kendall (KENDALL, 1984) systems are two possibilities. Additionally, the term *pre-shape* means the landmark configuration after extracting location and scale. According to Dryden and Mardia (2016), the shape is an equivalence class and, from a practical probabilistic perspective, it is common to work in the pre-shape space. To determine our action field in this thesis, we assume that under-study objects are in the plan (although would be in three or high dimensions), called as *planar-shape*.

According to Dryden and Mardia (2016), some SSA goals are: (i) to study how shape changes during growth; (ii) how shape changes during evolution; (iii) how shape is related to size; and (iv) how to describe shape variability. In order to model changes in shape it is necessary to define suitable statistical models according to Kent (1997). Among the various probability models used in shape data, two of the main distributions are: the complex Bingham ($\mathbb{C}B$) model proposed by Kent (1994) and, as one of its particular case, the complex Watson ($\mathbb{C}W$) model pioneered by Mardia and Dryden (1999). Some of their theoretical and computational properties have been developed in Dryden and Mardia (2016). Two reasons for wide demand of these models are their analytically tractable densities and by their two parameters are aligned with two important geometrical features in spherical data: mean shape and concentration degree. Biological hypothesis have been often formulating in the SSA framework, on which resulting data are in a particular Hilbert space. Tailored tests are then sought to take accurate decision in this area. Considering the $\mathbb{C}B$ and $\mathbb{C}W$ models, hypothesis tests have been proposed to check difference of mean shapes in two or multiple samples. Amaral, Dryden and Wood (2007) have provided a novel bootstrap hypothesis test based on the lambda statistics for directional and planar-shape data. Micheas and Dey (2005a) have proposed new Bayesian method based on the $\mathbb{C}W$

distribution that is used in detecting shape differences between two groups of mice. A common assumption which is made when formulating hypothesis tests for mean shape in the unit sphere is "samples should be strongly concentrated". This fact is geometrically justified, since models under this assumption are asymptotically equivalent to the ones in the tangent plan (commonly more tractable). However, sometime we find data having low concentration. This thesis addresses a way by means of SIT to treat with shape change on both high and low concentrations, assuming $\mathbb{C}B$ and $\mathbb{C}W$ probability distributions for shape data.

According to Salicrú et al. (1994) SIT measures (in particular stochastic divergence and entropy measures) may play an important role in statistical theory; specifically into estimation and hypothesis inference procedures under large samples. In this context, several works have been developed. Basseville (2013) has provided an annotated bibliography for investigations based on divergence measures for statistical data processing and inference problems. Nascimento, Cintra and Frery (2010) and Frery, Nascimento and Cintra (2014) have derived statistical tests based on stochastic distances for speckled intensity and polarimetric synthetic aperture radar (SAR) data, respectively. Frery, Cintra and Nascimento (2013) have furnished new entropy-based (for the Shannon, Rényi and restricted Tsallis entropy) hypothesis tests under the scaled complex Wishart law for polarimetric SAR data. In general this thesis builds statistical mechanisms to apply SIT to SSA in order to solve biological issues. New planar-shape tools to check homogeneity, mean shape and entropy-based variability in spherical samples are given. As first essays, we develop new distance-based two-sample tests for mean shapes of triangles (two-dimensional objects with three landmarks) without condition about concentration. To reach this goal, we assume that pre-shape data are well described by the $\mathbb{C}W$ distribution. We derive expressions for the Rényi, Kullback-Leibler, Bhattacharyya and Hellinger stochastic distances on the $\mathbb{C}W$ distribution.To quantify the performance of the proposed tests in contrast with the F2 literature test for mean shapes Mardia and Dryden (1999), both synthetic and real studies are made.

Posteriorly, as extensions to the first proposal, we provide new distance-based two-sample hypothesis tests (for both homogeneity and mean shape) for the $\mathbb{C}B$ distribution and number of landmarks higher than three. We derive the Rényi and Kullback-Leibler

divergences and the Bhattacharyya and Hellinger distances for the $\mathbb{C}B$ distribution. We prove also that the Kullback-Leibler discrepancy for the $\mathbb{C}B$ model is rotation invariant. In order to evaluate and compare our proposals with other four SSA mean shape tests (based on Hotelling T^2, Goodall, James and lambda statistics (DRYDEN; MARDIA, 2016)), a simulation study is also made to evaluate asymptotic and robustness properties. An application in evolutionary biology is also made. Finally, as third contribution, we tackle the proposal of new entropy-based multi-sample tests for variability in planar-shape. We develop closed-form expressions for the Rényi and Shannon entropies at the $\mathbb{C}B$ and $\mathbb{C}W$ models. From these quantities, hypothesis tests are obtained to assess if multiple spherical samples have the same degree of disorder. An application to real data using the second thoracic vertebra T2 of mice is done to assess possible effects of body weight on the shape of mouse vertebra.

1.1.1 Main contributions

In general, this thesis advances at proposing new tests, for medical and biological hypothesis using tools of SIT applied to SSA. As specific contributions, we propose:

i) derivation of expressions for Rényi, Bhattacharyya, Hellinger and Kullback-Leibler divergence measures under the $\mathbb{C}B$ and $\mathbb{C}W$ distributions;

ii) proposals of pivotal statistics for two-sample tests based on distances of item (i) for both triangle and planar-shape data;

iii) derivation of expressions for the Shannon and Rényi entropies under $\mathbb{C}B$ and $\mathbb{C}W$ models;

iv) proposals of two multi-sample pivotal statistics based on item (iii) entropies;

v) Monte Carlo experiment which illustrate how to use proposed and literature tests as well as compare them; discussions of applications on which proposed tests are successfully employed.

The main contributions of this thesis results in the submission of the following three papers.

Paper I: entitled by *Distance-based Hypothesis Tests for Triangle Shapes* submitted to the "Journal Computations and Graphical Statistics".

Paper II: entitled by *Divergence-based Pivotal Statistics for Planar-Shape* submitted to the "Journal of Multivariate Analysis". Status: under review.

Paper III: entitled by *Entropy-based Pivotal Statistics for Multi-Sample Problems in Planar-Shape* submitted to the "Test". Status: under review.

1.2 Computational Support

This manuscript has been written using the LaTeX software and its references were made in the BibTeX . Additional information about this typography system can be found in Syropoulos, Tsolomitis and Sofroniou (2003) and Grätzer (2007). As computational support, it has been used the R (Version of this program is freely available at <https://cran.r-project.org/>). statistical software. Main R packages and functions that we used in this thesis are furnished in Table 1.

Table 1 – R packages and functions that we used in this thesis

Packages	Functions
shapes	resampletest(), procrustes2d(), plotshapes(), procGPA(), riemdist(), realtocomplex(), preshape(), complextoreal(),
maxLik	maxBFGS()
optimx	optimx()
MASS	truehist(), ginv()
graphics	curve(), lines(), abline(), points(), plot(), par(),
hypergeo	genhypergeo()
grDevices	x11()
utils	data()
stats	runif(), pchisq(), qchisq(), ks.test(), optimize()
base	set.seed(), length(), mean(), var(), summary() , round(), matrix(), array(), function(), return(), sqrt( ), cbind(), return(), Conj(), sort( ), ceiling( ), print( ), source( ), qqplot( ), qchisq( ), pf( ), numeric(), factorial(),seq(), list(), eigen(), sum(),Re(),Im(), diag(), require(),library(),abs(),sin(),cos(),exp(), log() , sample(), t(), c( ),pf(), qf(),

1.3 Structure of the thesis

The thesis unfolds as follows. In the first chapter, we present motivation and goals of this thesis. The second chapter contains a brief survey about some concepts and preliminaries results which we use in this work. The third chapter describes the hypothesis testing methodology which is based on the divergence class proposed by Salicrú et al. (1994) for $\mathbb{C}W$ distributed data from triangles in the plan. The fourth chapter presents divergence-based two-sample hypothesis tests on the $\mathbb{C}B$ distribution for planar-shape. The fifth chapter provides entropy-based multi-sample hypothesis tests for in planar-shape on the $\mathbb{C}B$ distribution. Subsequently, the sixth chapter addresses the main conclusions. Future works are also presented. Finally, the Appendix chapter tackles in detail the derivations which are made during this thesis proposal.

2 BACKGROUND ABOUT STATISTICAL SHAPE ANALYSIS AND INFORMATION THEORY

2.1 Statistical Shape Analysis

The main concept of Statistical Shape Analysis (SSA) (KENT, 1995) is the meaning of shape, which is defined how what remains after removing location, scale and rotation. According to Dryden and Mardia (2016), the SSA is concerned with methodology for analyzing shapes in the presence of randomness. The under-study objects could be sampled at random from a population and the main aims of SSA are to estimate population mean shapes, to estimate the structure of population shape variability and to carry out inference on population features.

One way to describe a shape is to indicate a finite set of points in the boundary of the object. This finite number of points of each object are called *landmarks*. It is the main font of data for the description of shapes. Dryden and Mardia (2016) have furnished a precise definition for landmarks as follows.

Definition 2.1. *A landmark is a point of correspondence on each object that matches between and within populations.*

There are three basic types of landmarks : scientific, mathematical and pseudo-landmarks.

a) **Scientific landmark**: is a point assigned by an expert that corresponds between objects in some scientifically meaningful way; for example, the corner of an eye or the meeting of two sutures on a skull. In biological applications, such landmarks are also known as **anatomical landmarks**, and they designate parts of an organism that correspond in terms of biological derivation.

b) **Mathematical landmarks** are points located on an object according to some mathematical or geometrical properties of the figure; for instance, at a point of high curvature or at an extreme point;

c) **Pseudo-landmarks** are points on an object, located either around the outline or in between scientific or mathematical landmarks. Its use is to visualize a shape and not for analyzing it.

In this thesis, in our applications will take into consideration both mathematical and anatomical landmarks. The mathematical representation of landmarks in an object is called configuration.

Definition 2.2. *A configuration is the set of landmarks on a particular object. A configuration matrix* $\boldsymbol{X}$ *is a* $k \times m$ *matrix of cartesian coordinates of the* $k+1$ *landmarks in* m *dimensions. A configuration space is a space of all possible landmark coordinates.*

In this thesis, we consider $k \geq 2$ landmarks number (triangle and others) in $m = 2$ dimensions, called of *planar-shape*. Thus, the configuration matrix is

$$\boldsymbol{X} = [x_1 \; x_2]^\top = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \vdots & \vdots \\ x_{k+1,1} & x_{k+1,2} \end{pmatrix},$$

where x_j^1 and x_j^2 indicate coordinates in real and imaginary axes, respectively.

Some transformations must be made in $\boldsymbol{X}$ to remove the effects of location, scale and rotation. For $m = 2$, the mathematical configuration must be rewritten as a complex vector. Define a $(k+1) \times 1$ complex vector

$$\boldsymbol{z}^0 = [z_1^0, z_2^0, \ldots, z_{(k+1)}^0]^\top = [x_{1,1} + \mathrm{i}x_{1,2}, \ldots, x_{(k+1),2} + \mathrm{i}x_{(k+1),2}]^\top \text{ for } \mathrm{i} = \sqrt{-1},$$

which corresponds to the complex coordinates for landmarks.

To describe the shape of an object, it is necessary to specify a suitable coordinate system; i.e, a system that is invariant under translation, scaling and rotation of the configuration. There are several coordinate systems with this properties in SSA; as, for example, Bookstein coordinates for planar and triangular shapes Bookstein (1984), Bookstein (1986).

Watson (1986) have considered a coordinate system for triangle shapes. Kent (1994) has proposed Kent's polar coordinates in the tangent space (the tangent space is the linearized version of the shape space in the vicinity of a particular point of shape space).

Kendall (1984) has provided a coordinate systems similar to Bookstein's coordinates, but location is removed in a different manner. Rohlf (2000a) has compared Kendall's shape space, Kendall tangent space and the shape space implied by differences in angles of lines connecting pairs of landmarks. In this thesis, we will use the **Kendall coordinate** systems. To get these coordinates, we need to define the Helmert submatrix, as follows.

According to Dryden and Mardia (2016), the Helmert submatrix (denoted $\boldsymbol{H}$) is a $k \times (k+1)$ Helmert matrix without the first row. The full Helmert matrix $\boldsymbol{H}^F$ is a square $(k+1) \times (k+1)$ orthogonal matrix with its first row of elements equal to $1/\sqrt{k+1}$ and the remaining rows are orthogonal to the first row. We drop the first row of $\boldsymbol{H}^F$ so that the transformation $\boldsymbol{HX}$ does not depend on the original location of the configuration.

Definition 2.3. *The jth row of the Helmert submatrix $\boldsymbol{H}$ is given by*

$$(h_j, \ldots, h_j, -jh_j, 0, \ldots, 0) \quad \textit{for} \quad h_j = -[j(j+1)]^{-1/2}.$$

So the jth row of $\boldsymbol{H}$ consists of h_j repeated j times, followed by $-jh_j$ and $k-j$ zeros for $j = 1, \ldots, k$.

The next example illustrates $\boldsymbol{H}^F$ and $\boldsymbol{H}$.

Example 2.1. *For $k+1 = 3$ the full Helmert matrix is explicit by*

$$\boldsymbol{H}^F = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{pmatrix},$$

and the Helmert submatrix is

$$\boldsymbol{H} = \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{pmatrix}.$$

To remove the location effect in $\boldsymbol{z}^0 = [z_1^0, \ldots, z_{(k+1)}^0]^\top$, we pre-multiply it by the Helmert submatrix $\boldsymbol{H}$; i.e.,

$$\boldsymbol{z_H} = \boldsymbol{H}\boldsymbol{z}^0.$$

The origin is removed because coincident landmarks are not allowed. We refer to $\boldsymbol{z_H}$ as the **Helmertized landmark** coordinates. The **centred landmark** coordinates are an alternative choice for removing location and are given by:

$$\boldsymbol{z_C} = \boldsymbol{C}\boldsymbol{z}^0 \quad \text{with} \quad \boldsymbol{C} = \boldsymbol{I}_{(k+1)} - \frac{1}{(k+1)} 1_{(k+1)} 1_{(k+1)}^\top,$$

where $\boldsymbol{I}_{(k+1)}$ is the identity matrix of order $(k+1) \times (k+1)$ and $1_{(k+1)}$ is an ones vector. We can revert back to the centred landmark coordinates from the Helmertized landmark coordinates by pre-multiplying by $\boldsymbol{H}^\top$. Thus

$$\boldsymbol{H}^\top \boldsymbol{z_H} = \boldsymbol{H}^\top \boldsymbol{z_H} = \boldsymbol{H}^\top \boldsymbol{H} \boldsymbol{z}^0 = \boldsymbol{C}\boldsymbol{z}^0.$$

Then, $\boldsymbol{z_C}$ are the centred landmarks and $\boldsymbol{C}$ is the centring matrix.

Example 2.2. *Consider one particular mathematical configuration of an individual which was obtained of male chimpanzee skull data having $k+1 = 8$ landmarks. The full dataset has already been described in O'Higgins and Dryden (1993) and Dryden and Mardia (2016). The original complex landmarks of this data are*

$$\boldsymbol{z}^0 = [43 - 21\mathrm{i}, 0 + 0\mathrm{i}, 0 + 34\mathrm{i}, 14 + 101\mathrm{i}, 25 + 179\mathrm{i}, 40 + 150\mathrm{i}, 75 + 104\mathrm{i}, 90 + 31\mathrm{i}]^\top.$$

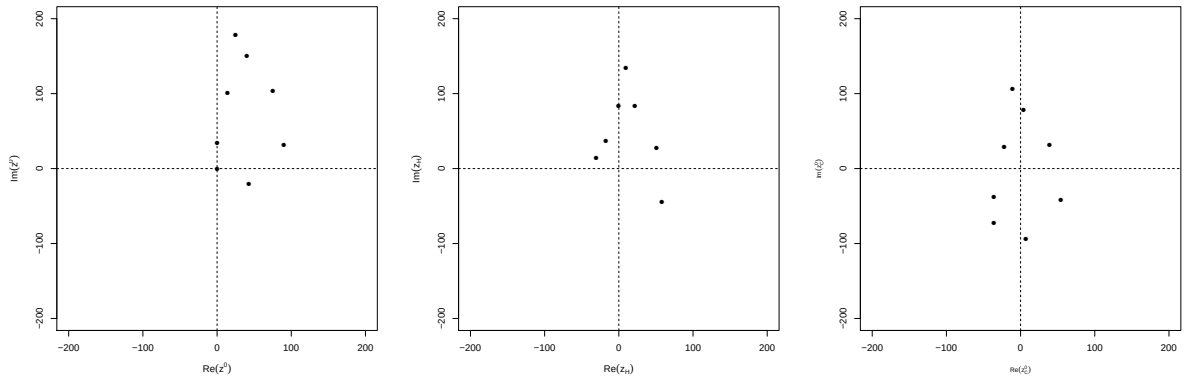
The Helmertized landmarks are

$$\begin{aligned}\boldsymbol{z_H} = &[-30.41 + 14.85\mathrm{i}, -17.55 + 36.33\mathrm{i}, -0.29 + 83.72\mathrm{i}, 9.62 + 134.61\mathrm{i},\\ &21.54 + 83.44\mathrm{i}, 50.61 + 27.93\mathrm{i}, 57.86 - 44.10\mathrm{i}]^\top.\end{aligned}$$

The centred landmarks are

$$\begin{aligned}\boldsymbol{z}_{\boldsymbol{C}}^0 = &[7.13 - 93.25\mathrm{i}, -35.87 - 72.25\mathrm{i}, -35.88 - 38.25\mathrm{i}, -21.87 + 28.75\mathrm{i},\\ &-10.87 + 106.75\mathrm{i}, 4.13 + 77.75\mathrm{i}, 39.12 + 31.75\mathrm{i}, 54.12 - 41.25\mathrm{i}]^\top.\end{aligned}$$

The Figure 1 represents the original, helmertized and centered settings for a male chimpanzee skull data. The dataset was assessed by command `panm.dat` *by means of platform* `R` *in the package* `shapes`*.*



(a) Original configuration (b) Helmertized configuration (c) Centred configuration

Figure 1 – Illustration of original, helmertized and centered configurations of a male chimpanzee skull. The full dataset has already been described in Dryden and Mardia (2016, pp. 19).

To remove the scale effect, we divided the obtained Helmertized configuration ($\boldsymbol{z_H}$) by its norm

$$\boldsymbol{z} = \frac{\boldsymbol{z_H}}{||\boldsymbol{z_H}||} = \frac{\boldsymbol{z_H}}{\sqrt{\boldsymbol{z_H}^* \boldsymbol{z_H}}},$$

where $(\cdot)^*$ represents the complex conjugate operator and $||\cdot||$ denotes the norm of a complex vector. According to Kendall (1984), this normalized vector $\boldsymbol{z}$ is called pre-shape, defined as follows.

Definition 2.4. *The* ***pre-shape*** *of a configuration matrix* $\boldsymbol{X}$ *is given by*

$$\boldsymbol{Z} = \frac{\boldsymbol{X_H}}{||\boldsymbol{X_H}||} = \frac{\boldsymbol{HX}}{||\boldsymbol{HX}||},$$

which is invariant under the location and scaling of the original configuration.

Example 2.3. *Considering the same dataset of Example 2.2, the pre-shape of a mathematic configuration individual of the male chimpanzee skull is given by*

$$\begin{aligned}\boldsymbol{z} = [&-0.14 + 0.07\text{i}, -0.08 + 0.17\text{i}, 0.00 + 0.40\text{i},\\ &0.05 + 0.64\text{i}, 0.10 + 0.40\text{i}, 0.24 + 0.13\text{i}, 0.28 - 0.21\text{i}]^\top . \end{aligned} \tag{2.1}$$

The Figure 2 draws the pre-shape of the configuration individual 2.1.

The pre-shape space denoted by $\boldsymbol{S}_m^k$ (k landmarks and m dimensions) is the space of all possible pre-shapes. Formally Dryden and Mardia (2016) define as follows.

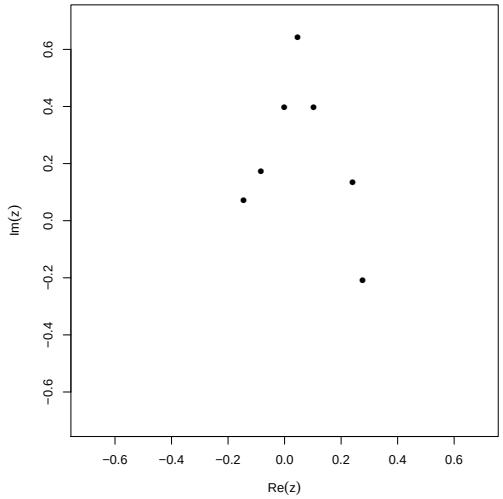


Figure 2 – Illustration of the pre-shape of a configuration individual from male chimpanzee skull data.

Definition 2.5. *The **pre-shape space** $\boldsymbol{S}_m^k$ is the orbit space of the non coincident k point set configurations in $\mathbb{R}^m$ under the action of translation and isotropic scaling. This space collapses into a complex hypersphere of dimension k, that is,*

$$\mathbb{C}\boldsymbol{S}^{k-1} = \{\boldsymbol{z} \in \mathbb{C}^k; ||\boldsymbol{z}|| = 1\},$$

where $\mathbb{C}^k$ is the complex space of dimension k.

Note that the term "pre-shape" means that we are one step away from shape: rotation still has to be removed.

In order to remove rotation information from the configuration, it identifies all rotated versions of the pre-shape with each other and this set or equivalence class is the shape of $\boldsymbol{X}$.

Definition 2.6. *The **shape** of a configuration matrix $\boldsymbol{X}$ is all the geometrical information about $\boldsymbol{X}$ that is invariant under location, rotation and isotropic scaling (Euclidean similarity transformations). The shape can be represented by the equivalence class given by*

$$[\boldsymbol{X}] = \{\boldsymbol{Z}\boldsymbol{\Gamma}; \boldsymbol{\Gamma} \in SO(m)\},$$

where $SO(m)$ is the special orthogonal group of rotations and $\boldsymbol{Z}$ is the pre-shape of $\boldsymbol{X}$.

For $m = 2$, in the planar-shape case, the shape of an original configuration is given by

$$[\boldsymbol{z}] = \{\boldsymbol{z}\exp(\mathrm{i}\theta); \theta \in [0, 2\pi)\}.$$

Definition 2.7. *The **shape space** is the space of all shapes. Formally, the shape space denoted by $\boldsymbol{\Sigma}_m^k$ is the orbit space of the non coincident k point set configurations in $\mathbb{R}^m$ under the action of the Euclidean similarity transformations (translation, rotation and scale).*

The shape of $\boldsymbol{X}$ is a set: an equivalence class under the action of the group of similarity transformations. In order to visualize shapes it is often convenient to choose a particular member of the shape set $[\boldsymbol{X}]$, which refers to the following concept.

Definition 2.8. *An **icon** is a particular member of the shape set $[\boldsymbol{X}]$ which is taken as being representative of the shape.*

According to Dryden and Mardia (2016), the word icon can mean "image or likeness" and it is appropriate as we use the icon to picture a representative figure from the shape equivalence class which has a resemblance to the other members, that is the objects of the class are all similar. The centred pre-shape is a suitable choice of icon.

Example 2.4. *Referring again to dataset of the Example 2.1, the icon of a configuration individual of the male chimpanzee skull data is*

$$\begin{aligned}\boldsymbol{z_C} = [&0.03 - 0.44\mathrm{i}, -0.17 - 0.34\mathrm{i}, -0.17 - 0.18\mathrm{i}, -0.10 + 0.14\mathrm{i}, -0.05 + 0.51\mathrm{i}\\ &0.02 + 0.37\mathrm{i}, 0.19 + 0.15\mathrm{i}, 0.26 - 0.20\mathrm{i}\,]^\top .\end{aligned}$$

In the Figure 3, we can see the illustration of the centred pre-shape icon, the rotation is unchanged from the original configuration $\boldsymbol{z}^0$.

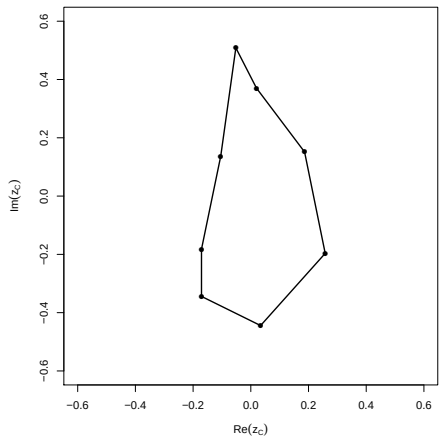


Figure 3 – An icon for an individual male chimpanzee skull data, which is the centred pre-shape.

Faced with a variety of terminology presented until here, we display in the Figure 4 (adapted from Dryden and Mardia (2016)) a short resume about the hierarchies in different spaces. In this diagram, we show the steps for obtaining the shape from one of an original configuration.

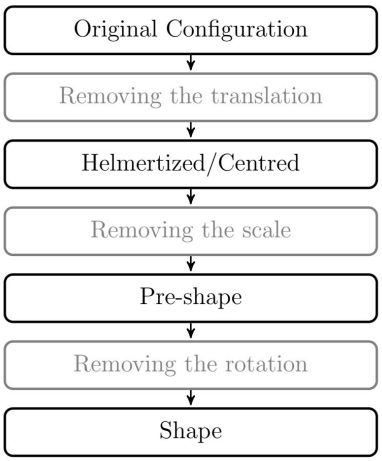


Figure 4 – The hierarchies of the various spaces.

Another very important coordinate system in SSA and widely used in the later chapters is the **Kent's polar pre-shape coordinates**. Kent (1994) has proposed some non standard polar coordinates on the pre-shape sphere for two-dimensional objects. Given a point $[z_1, z_2, \ldots, z_k]^\top$ on $\mathbb{C}\boldsymbol{S}^{k-1}$, we transform it to $[s_1, \ldots, s_{k-1}, \theta_1, \ldots, \theta_k]$, where

$$\mathrm{Re}(z_j) = s_j^{1/2}\cos(\theta_j) \text{ and } \mathrm{Im}(z_j) = s_j^{1/2}\sin(\theta_j),$$

for $j = 1, \ldots, k$, $s_j \geq 0$, $\theta \in [0, 2\pi)$ and $s_k = 1 - \sum_{i=1}^{k-1} s_j$. The coordinates $s_1, s_2, \ldots, s_{k-1}$ are on the $(k-1)$-dimensional unit simplex in $\mathbb{R}^{k-1}$ defined by

$$\mathcal{S}_{k-1} = \left\{ [s_1, \ldots, s_{k-1}]^\top;\ s_j \geq 0 \quad \text{and} \quad \sum_{j=1}^{k-1} s_j \leq 1 \right\}.$$

The vector $\mathbf{s} = [s_1, \ldots, s_{k-1}]^\top$ follows the truncated multivariate exponential (TME) distribution to a simplex with density given by Kent, Constable and Er (2004)

$$f_{TE}(\boldsymbol{s}) = b(\lambda)\prod_{j=1}^{k-1}\lambda_j \exp(-\lambda_j s_j), \quad s_j \geq 0, \quad j = 1, \ldots k-1,$$

where $b(\lambda) = \left\{ \prod_{j=1}^{k-1}[1 - \exp(-\lambda_j)] \right\}^{-1}$. By identifying the complex pre-shape sphere with $\mathcal{S}_{k-1} \times [0, 2\pi)^k$, we have the volume measure of $\mathbb{C}\boldsymbol{S}^{k-1}$ as:

$$2^{1-k} ds_1 \ldots ds_{k-1} d\theta_1 \ldots d\theta_k.$$

The total volume is $2\pi^k/(k-1)!$, since the volume of the j-dimensional simplex (DRYDEN; MARDIA, 2016) is $1/j!$, for $j = 1, 2, 3, \ldots$.

2.1.1 Full Procrustes coordinates and mean shape

The expression "Procrustes analysis" was first used by Hurley and Cattell (1962) in factor analysis. In Greek mythology, Procrustes were the nickname of a robber Damastes, who lived by the road from Eleusis to Athens. He would offer travellers a room for the night and fit them to the bed; by stretching them. If they were too short or chopping off their limbs if they were too tall. The analogy is rather tenuous but we can regard one configuration as the bed and the other as the person being "translated", "rotated" and possibly "rescaled" so as to fit as close as possible to the bed.

The goal of full Procrustes coordinates is to match the shape of two objects. The idea is to be able to compare the shapes between two objects independently of their rotation and scale. Consider the centred unit size configurations $\boldsymbol{y} = [y_1, \ldots, y_k]^\top$ and $\boldsymbol{w} = [w_1, \ldots, w_k]^\top$ where $||\boldsymbol{y}|| = ||\boldsymbol{w}|| = 1$, $\boldsymbol{y}^* 1_k = \boldsymbol{w}^* 1_k = 0$, $(\cdot)^*$ represents the complex conjugate operator and 1_k is the $k \times 1$ vector of ones. In order to compare the configurations in shape we need to establish a measure of distance between the two shapes. A suitable procedure is to match $\boldsymbol{w}$ to $\boldsymbol{y}$ using the similarity transformations. The differences between the fitted and observed indicate the magnitude of the difference in shape between $\boldsymbol{w}$ and $\boldsymbol{y}$. Consider the *complex regression* equation

$$\boldsymbol{y} = (a + \mathrm{i}b)1_k + \varpi \exp(\mathrm{i}\theta)\boldsymbol{w} + \boldsymbol{\epsilon} = [1_k, \boldsymbol{w}]\boldsymbol{A} + \boldsymbol{\epsilon} = \boldsymbol{X}_D\boldsymbol{A} + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]^\top = [(a + \mathrm{i}b)1_k, \varpi \exp(\mathrm{i}\theta)]$ are the 2×1 complex parameters with translation a $a + \mathrm{i}b$, scale $\varpi > 0$ and rotation $0 \leq \theta < 2\pi$; $\boldsymbol{\epsilon}$ is a $k \times 1$ complex error vector; and $\boldsymbol{X}_D = [1_k, \boldsymbol{w}]$ is the $k \times 2$ *design matrix*. To carry out the superimposition we could estimate $\boldsymbol{A}$ by minimizing the least squares objective function, the sum of square errors

$$D^2(\boldsymbol{y}, \boldsymbol{w}) = \boldsymbol{\epsilon}^*\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{A})^*(\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{A}).$$

From this expression (2.2), we define the full Procrustes fit (registration) of $\boldsymbol{w}$ onto $\boldsymbol{y}$.

Definition 2.9. *Let $\boldsymbol{w}$ and $\boldsymbol{y}$ in $\mathbb{C}^k$, such that $\boldsymbol{y}^* 1_k = \boldsymbol{w}^* 1_k = 0$. The full Procrustes fit of $\boldsymbol{w}$ onto $\boldsymbol{y}$ is*

$$\boldsymbol{w}^P = \boldsymbol{X}_D\widehat{\boldsymbol{A}} = (\widehat{a} + \mathrm{i}\widehat{b})1_k + \widehat{\varpi} \exp(\mathrm{i}\theta)\boldsymbol{w},$$

where $\widehat{\boldsymbol{A}}$ is obtained by the expression

$$\arg\min_{a,b,\varpi,\theta} (\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{A})^*(\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{A}) = \arg\min_{a,b,\varpi,\theta} ||\boldsymbol{y} - (a + \mathrm{i}b)1_k - \varpi \exp(\mathrm{i}\theta)\boldsymbol{w}||^2.$$

It follows from Dryden and Mardia (2016) that full Procrustes fit has matching parameters

$$(\widehat{a} + \mathrm{i}\widehat{b}) = 0;\ \widehat{\theta} = \arg(\boldsymbol{w}^*\boldsymbol{y}) = -\arg(\boldsymbol{y}^*\boldsymbol{w})) \text{ and } \widehat{\varpi} = ||\boldsymbol{w}^*\boldsymbol{y}||/||\boldsymbol{w}||^2, \tag{2.3}$$

where $\arg(\cdot)$ is the argument of a complex number.

Thus, from Definition 2.9 and Equations in (2.3), we have the full Procrustes fit of $\boldsymbol{w}$ onto $\boldsymbol{y}$ is given by

$$\begin{aligned}\boldsymbol{w}^P = \boldsymbol{X}_D\widehat{\boldsymbol{A}} &= (\widehat{a} + \mathrm{i}\widehat{b})1_k + \widehat{\varpi}\exp(\mathrm{i}\theta)\boldsymbol{w} = \frac{||\boldsymbol{w}^*\boldsymbol{y}||}{||\boldsymbol{w}||^2}\exp[\mathrm{i}\arg(\boldsymbol{w}^*\boldsymbol{y})]\boldsymbol{w} \\ &= \frac{||\boldsymbol{w}^*\boldsymbol{y}||}{||\boldsymbol{w}||^2}\frac{\boldsymbol{w}^*\boldsymbol{y}}{||\boldsymbol{w}^*\boldsymbol{y}||}\boldsymbol{w} = \frac{\boldsymbol{w}^*\boldsymbol{y}}{||\boldsymbol{w}||^2}\boldsymbol{w}.\end{aligned}$$

In order to define the full Procrustes distance between two shapes, we use the residual vector given by

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}_D\widehat{\boldsymbol{A}} = [\boldsymbol{I}_k - \boldsymbol{X}_D(\boldsymbol{X}_D^*\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^*]\boldsymbol{y}.$$

Therefore, we will to compare shapes by means of $\boldsymbol{w}^P$ and $\boldsymbol{y}$,

$$d_F(\boldsymbol{y}, \boldsymbol{w}) = \sqrt{\boldsymbol{r}^*\boldsymbol{r}} = \sqrt{\boldsymbol{y}^*\boldsymbol{y} - (\boldsymbol{y}^*\boldsymbol{w}\boldsymbol{w}^*\boldsymbol{y})/(\boldsymbol{w}^*\boldsymbol{w})}.$$

Now this expression is not symmetric in $\boldsymbol{y}$ and $\boldsymbol{w}$ unless $\boldsymbol{y}^*\boldsymbol{y} = \boldsymbol{w}^*\boldsymbol{w}$. A convenient standardization is to take the configurations to be unit size, that is $||\boldsymbol{y}||^2 = ||\boldsymbol{w}||^2 = 1$. So, if we include standardization, then we obtain a suitable measure of shape distance.

Definition 2.10. *The* ***full Procrustes distance*** *between complex configuration* $\boldsymbol{w}$ *and* $\boldsymbol{y}$ *is expressed by*

$$\begin{aligned}d_F(\boldsymbol{y}, \boldsymbol{w}) &= \inf_{a,b,\varpi,\theta}\left\|\frac{\boldsymbol{y}}{||\boldsymbol{y}||} - \frac{\boldsymbol{w}}{||\boldsymbol{w}||}\varpi\exp(\mathrm{i}\theta) - 1_k(a + \mathrm{i}b)\right\| \\ &= \left(1 - \frac{\boldsymbol{y}^*\boldsymbol{w}\boldsymbol{w}^*\boldsymbol{y}}{\boldsymbol{w}^*\boldsymbol{w}\boldsymbol{y}^*\boldsymbol{y}}\right)^{1/2}.\end{aligned}$$

The full Procrustes distance is natural from a statistical point of view, obtained from a least squares criterion and optimizing over the full set of similarity parameters. The squared full Procrustes distance naturally appears exponentiated in the density for many simple probability distributions for shape (DRYDEN; MARDIA, 2016). However, this is not the only choice of distance between shapes. Since the shapes of configurations are

represented by fibres (fibres on the pre-shape sphere correspond one to one with shapes in the shape space, and so we can think of a fibre as representing the shape of a configuration) on the pre-shape sphere, we can define the distance between two shapes as the closest great circle distance between the fibres on the pre-shape sphere. This shape distance is called the **Procrustes (Riemannian) distance** in shape space and is denoted by ρ. Thus, the angle between the complex pre-shapes $\boldsymbol{y}$ and $\boldsymbol{w}$ is defined in Dryden and Mardia (2016) by

$$\rho(\boldsymbol{w}, \boldsymbol{y}) = \arccos(|\boldsymbol{y}^* \boldsymbol{w}|), \text{ where } 0 \leq \rho(\boldsymbol{w}, \boldsymbol{y}) \leq \pi/2.$$

Note that, $\rho(\cdot, \cdot)$ can be considered as the smallest angle (with respect to rotations of the pre-shapes) between the vectors corresponding to $\boldsymbol{y}$ and $\boldsymbol{w}$ on the pre-shape sphere. There exist a relationship between the Procrustes distance and full Procrustes distance, expressed by

$$\cos(\rho(\boldsymbol{w}, \boldsymbol{y})) = (1 - d_F^2(\boldsymbol{w}, \boldsymbol{y}))^{1/2}.$$

The shape distance allows comparing two configurations, but comparing several at the same time it is not possible. To solve this problem we present an important concept in SSA, extensively used in this thesis: the *mean shape* of a random sample of configurations.

According to Dryden et al. (2014), there are different notions of mean shape, and it is important to distinguish between them. The different means are obtained by using different distances, embeddings and projections in the definition of the mean. For example, Fréchet mean shape Le and Kume (2000), Procrustes mean shape Dryden (1991), extrinsic and intrinsic mean shape Bhattacharya, Patrangenaru et al. (2003) and Bhattacharya and Patrangenaru (2005) among others. In this thesis, we use the concepts of full Procrustes mean shape (is an extrinsic mean for planar landmark data using the complex Veronese–Whitney embedding on pre-shape space) based in Dryden and Mardia (2016, chap. 3–4), as follows.

Definition 2.11. *Let $\boldsymbol{x}$ be a random quantity on pre-shape space. The population* ***full Procrustes mean shape*** *is given by*

$$\boldsymbol{\mu} = \arg\inf_{\boldsymbol{\mu}} \mathbb{E}[d_F(\boldsymbol{x}, \boldsymbol{\mu})].$$

Now, consider the situation where a random sample of centred configurations $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ is available and we wish to estimate a population mean shape, such as the population full Procrustes mean.

Definition 2.12. *The full Procrustes estimate of mean shape* $[\widehat{\boldsymbol{\mu}}]$ *is obtained by minimizing (over* $\boldsymbol{\mu}$*) the sum of square full Procrustes distances from each* $\boldsymbol{w}_i$ *to an unknown unit size mean configuration* $\boldsymbol{\mu}$*, that is*

$$[\widehat{\boldsymbol{\mu}}] = \arg\inf_{\boldsymbol{\mu}} \sum d_F^2(\boldsymbol{w}_i, \boldsymbol{\mu}).$$

Kent (1994) has furnished an result to obtain the full Procrustes estimate of mean shape by means of product matrix $\boldsymbol{S}$, as follows.

Proposition 2.1. *The full Procrustes mean shape* $[\widehat{\boldsymbol{\mu}}]$ *can be defined as the eigenvector corresponding to the largest eigenvalue of the complex sum of squares and product matrix*

$$\boldsymbol{S} = \sum_{j=1}^{n} z_j z_j^*,$$

where $z_j = \boldsymbol{w}_j / ||\boldsymbol{w}_j||$ *for* $j = 1, \ldots, n$ *are the pre-shapes.*

Thus, $\widehat{\boldsymbol{\mu}}$ is given by the complex eigenvector which is corresponding to the largest eigenvalue, or the dominant eigenvector of $\boldsymbol{S}$. The eigenvector is unique (up to a rotation - all rotations of $\widehat{\boldsymbol{\mu}}$ are also solutions, but all these correspond in the same way), provided that there is a single eigenvalue greater than $\boldsymbol{S}$. A relevant comment about the estimative of mean shape is made by Rohlf (2003), which has realized sampling experiments were performed to investigate mean square error and bias in estimates of mean shape produced by different geometric morphometric methods.

2.2 Shape Distributions

An extremely important step in SSA (like as to detect changes in shape) is to define a probability model to describe planar-shape data (DRYDEN; MARDIA, 2016). One of main distributions for planar-shape is the $\mathbb{C}B$ model. The real Bingham distribution was introduced by Bingham (1974) as a generalization of the Dimroth–Watson model Watson (1965). Its complex version of the Bingham distribution and some of its mathematical

properties have been derived by Kent (1994). One of its simulation procedures has been proposed by Kent, Constable and Er (2004). They have shown such algorithm depends only of the truncated multivariate exponential (TME) distribution. In this thesis, we use to produce synthetic experiment a truncated simplex method combined with the TME law.

Several works have been developed using the real Bingham distribution and its complex counterpart. Kent (1987) has proposed an asymptotic expansion for the Bingham normalization constant. Kume and Wood (2005) have furnished a saddlepoint approximation for the Bingham normalizing constant. Kume and Wood (2007) have shown that an arbitrary joint partial derivative of the Bingham normalizing constant may be proportional to the normalizing constant of a Bingham distribution of higher dimension. Amaral et al. (2010) have derived bootstrap procedures for constructing confidence regions for the mean shape of objects labelled by landmarks in two dimensions. Dryden et al. (2005) have developed a mathematical treatment for the statistical analysis of high-dimensional spherical and shape data. Dore et al. (2016) have provided a bias-corrected estimation method for the $\mathbb{C}B$ distribution. In the Bayesian context, Micheas, Dey and Mardia (2006) have proposed a maximum posteriori probability estimators having the $\mathbb{C}B$ model like the posterior distribution. Amaral, Floréz and Cysneiros (2013) have furnished methods to detect influential observations in samples of pre-shape under such model. Leu and Damien (2014) have developed full Bayesian analysis for planar landmark data arising from the $\mathbb{C}B$ distribution.

A particular case of the $\mathbb{C}B$ distribution is the complex Watson ($\mathbb{C}W$) model, proposed by Mardia and Dryden (1999). Maximum likelihood estimation and hypothesis testing procedures for the $\mathbb{C}W$ model in one and two samples are furnished. These models are adequate for shape analysis data according to Dryden and Mardia (2016). It is relevant to mention how the $\mathbb{C}W$ distribution has been explored in the literature. Micheas and Dey (2005a) have proposed a novel Bayesian method based on the $\mathbb{C}W$ distribution that is used in detecting shape differences between the second thoracic vertebrae for two groups of mice, small and large, categorized according to their body weight. Micheas, Dey and Mardia (2006) have developed maximum likelihood and Bayesian estimation methods to describe shape and obtain confidence bounds and credible regions for shapes.

2.2.1 The complex Bingham distribution

In this subsection, we briefly discuss a model we use in this work for describing pre-shape data, admitting each k-dimensional random vector is associated to an object in the plan with $k+1$ landmarks. The $\mathbb{C}B$ distribution is defined on the unit complex sphere in $\mathbb{C}^k$ for $k \geq 2$, say $\mathbb{C}\boldsymbol{S}^{k-1} = \{\boldsymbol{z} = [z_1, \ldots, z_k]^\top; \boldsymbol{z}^*\boldsymbol{z} = 1\}$, and its probability density function (pdf) is given by

$$f(\boldsymbol{z}; \boldsymbol{A}) = c(\boldsymbol{A})^{-1} \exp\left(\boldsymbol{z}^* \boldsymbol{A}\, \boldsymbol{z}\right), \quad \boldsymbol{z} \in \mathbb{C}\boldsymbol{S}^{k-1},$$

where $(\cdot)^*$ represents the complex conjugate operator, $\boldsymbol{A}$ is a $k \times k$ hermitian matrix (i.e. $\boldsymbol{A} = \boldsymbol{A}^*$) and $c(\boldsymbol{A})$ is the normalizing constant (DRYDEN; MARDIA, 2016, pp. 221) given by

$$c(\boldsymbol{A}) = 2\pi^k \sum_{j=1}^{k} a_j \exp\left(\lambda_j\right) \quad \text{with} \quad a_j^{-1} = \prod_{\substack{i=1 \\ i \neq j}}^{k} (\lambda_j - \lambda_i), \tag{2.4}$$

where $\lambda_1 < \lambda_2 < \cdots < \lambda_{k-1} < \lambda_k = 0$ denotes the eigenvalues of $\boldsymbol{A}$. This normalizing constant is such that $c(\boldsymbol{A}) = c(\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ and $\text{diag}(\cdot)$ denotes the diagonal matrix of a vector argument. This distribution is denote as $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$. The $\mathbb{C}B$ distribution has the property of invariance under scalar rotation; i.e., $\boldsymbol{z}$ and $\exp(\mathrm{i}\theta)\boldsymbol{z}$ have the same distribution for all $\theta \in [0, 2\pi)$, property commonly used to define shape. Dryden and Mardia (2016) have stated that if an object is rotated, then its density and statistical properties are similar to those due to object at the original rotation. According to Kent, Constable and Er (2004), the matrix parameters $\boldsymbol{A}$ and $\boldsymbol{A} + \Upsilon \boldsymbol{I}_k$, for $\Upsilon \in \mathbb{R}$, define the same distribution. Hence, without loss of generality, we may shift the eigenvalues of $\boldsymbol{A}$ so that they are non positive with the largest one equaling 0. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k = 0$ denote the eigenvalues of $-\boldsymbol{A}$; i.e. λ_j, here corresponds to $-\lambda_j$ in (2.4), with $j = 1, \ldots, k$.

Kent (1994) has affirmed that the $\mathbb{C}B$ distribution can be obtained by conditioning a zero mean complex multivariate normal distribution to have unit norm. In particular, if $\boldsymbol{\omega} \sim \mathbb{C}N_k(0, \boldsymbol{\Sigma})$, then $\boldsymbol{\omega} \mid \{||\boldsymbol{\omega}|| = 1\} \sim \mathbb{C}\boldsymbol{B}_{k-1}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1})$. To obtain the $\mathbb{C}B$ log-likelihood function, we consider $\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n$, for $n \geq k$, be a sample of $(k+1)$-dimensional unit complex vectors and

$$\boldsymbol{S} = \sum_{j=1}^{n} \boldsymbol{z}_j \boldsymbol{z}_j^* \tag{2.5}$$

be $k \times k$ complex matrix, termed by *product matrix*. Suppose that the eigenvalues of $\boldsymbol{S}$ are positive, distinct and denoted by $0 < l_1 < \ldots < l_k$ and let $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$ be the corresponding eigenvectors. According to Dryden and Mardia (2016), the log-likelihood function at $\boldsymbol{A}$ is given by

$$\ell(\boldsymbol{A}; \boldsymbol{z}) = \sum_{i=1}^{n} \boldsymbol{z}_i^* \boldsymbol{A} \boldsymbol{z}_i - n \log c(\boldsymbol{A}) = \sum_{j=1}^{k} \lambda_j \boldsymbol{\gamma}_j^* \boldsymbol{S} \boldsymbol{\gamma}_j - n \log c(\boldsymbol{\Lambda}).$$

Considering $\lambda_1 < \lambda_2 < \cdots < \lambda_{k-1} < \lambda_k = 0$ constants, the ML estimate of eigenvectors $(\boldsymbol{\gamma}_j)$ referring to matrix $\boldsymbol{A}$ are given by $\widehat{\boldsymbol{\gamma}}_j = \boldsymbol{g}_j$, for $j = 1, \ldots, k$. So,

$$\ell(\boldsymbol{A}; \boldsymbol{z}) = \sum_{j=1}^{k} l_j \lambda_j - n \log c(\boldsymbol{\Lambda}). \tag{2.6}$$

Deriving the expression (2.6), we find the system of non linear equations

$$\frac{\partial \ell(\boldsymbol{A}; \boldsymbol{z})}{\partial \lambda_j} = \frac{1}{n} l_j, \quad j = 1, \ldots, k-1.$$

There is not closed-form solution for ML estimators of eigenvalues. Thus, it is need to use interactive process Nocedal and Wright (2006). Dryden and Mardia (2016) have proved under high concentrations that

$$\widehat{\lambda}_j \cong -n/l_j \quad \text{for} \quad j = 1, \ldots, k-1.$$

Details about ML estimators can be found in Amaral, Floréz and Cysneiros (2013) and Dore et al. (2016).

2.2.2 The complex Watson distribution

The $\mathbb{C}W$ distribution is an especial case of the $\mathbb{C}B$ model, they are defined in the same support, namely, $\mathbb{C}\boldsymbol{S}^{k-1}$. The $\mathbb{C}W$ pdf (MARDIA, 1999) is defined by

$$f(\boldsymbol{z}; \boldsymbol{\mu}, \kappa) = c_1(\kappa)^{-1} \exp\{\kappa \cos^2[\rho(\boldsymbol{H}^\top \boldsymbol{z}, \boldsymbol{H}^\top \boldsymbol{\mu})]\}, \quad \boldsymbol{z} \in \mathbb{C}\boldsymbol{S}^{k-1},$$

with $\rho(\boldsymbol{H}^\top \boldsymbol{z}, \boldsymbol{H}^\top \boldsymbol{\mu}) = \cos^{-1}(|\boldsymbol{z}^* \boldsymbol{\mu}|)$, for $0 \leq \rho(\boldsymbol{H}^\top \boldsymbol{z}, \boldsymbol{H}^\top \boldsymbol{\mu}) \leq \pi/2$, and the function $|\cdot|$ represents a complex number norm. This distribution is denoted as $\mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}, \kappa)$, where κ is a concentration parameter, $\boldsymbol{\mu}$ is the modal vector on pre-shape sphere and ρ is the Procrustes distance. The integrating constant defined in (**??**) is given by

$$c_1(\kappa) = \frac{2\pi^k}{(k-1)!} {}_1F_1(1, k, \kappa),$$

where

$$ {}_1F_1(a,b,x) = 1 + \frac{a}{b}\frac{x}{1!} + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x^3}{3!} + \ldots, \quad \text{with} \quad a,b,x \in \mathbb{R}, $$

is the confluent hypergeometric function (ABRAMOWITZ; STEGUN, 1964). The concentration parameter (κ) can assume negative values, but the assumption $\kappa > 0$ is the most common in shape analysis. For positive concentration, the modal pre-shape direction is $\boldsymbol{\mu}\exp(\mathrm{i}\theta)$, where $\theta \in [0, 2\pi)$ is an arbitrary rotation angle. If $\kappa = 0$, one gets the uniform distribution in the complex unitary sphere. Commonly, we consider the case of large κ. Second Mardia and Dryden (1999), if κ is large, then the $\mathbb{C}W$ model tends to a complex multivariate normal distribution in the tangent space. The $\mathbb{C}W$ model has the property of invariant under scalar rotation according to Kent (1994) and Mardia and Dryden (1999). This characteristic is very relevant in SSA, because if an object is rotated, then its density and statistical properties are similar to those due to object at the original rotation.

The ML estimation on $\mathbb{C}W$ distribution presented below was found in Dryden and Mardia (2016). Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be a random sample from a population modelled by the $\mathbb{C}W$ distribution, with $n \geq k$ and consider the product matrix defined in (2.5). According to Mardia and Dryden (1999) the log-likelihood function at $(\boldsymbol{\mu}, \kappa)$ is given by

$$ \ell(\boldsymbol{\mu}, \kappa; \boldsymbol{z}) = \kappa \operatorname{tr}\left(\sum_{j=1}^{k} l_j \boldsymbol{g}_j \boldsymbol{g}_j^*\right) - n\log(c_1(\kappa)). $$

Assuming κ constant, the ML estimator for $\boldsymbol{\mu}$ is

$$ \widehat{\boldsymbol{\mu}} = \exp(\mathrm{i}\tilde{\alpha})\boldsymbol{g}_k, $$

where $\tilde{\alpha}$ is a arbitrary rotation angle ($0 \leq \tilde{\alpha} < 2\pi$). The ML estimator for κ is determined by solving

$$ \frac{\partial[\log c_1(\kappa)]}{\partial\kappa} = \frac{1}{n} l_k. $$

Note there is not closed-form expressions, but under high concentration the following an approximation holds:

$$ \widehat{\kappa} \approx \frac{n(k-2)}{n - l_k}. $$

2.2.3 The relation between $\mathbb{C}B$ and $\mathbb{C}W$ models

According to Mardia and Dryden (1989b) the $\mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$ and $\mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}, \kappa)$ models are equivalent, that is, when there are just two distinct eigenvalues in $\boldsymbol{A}$ (a single distinct largest

eigenvalue and all other eigenvalues being equal). In this case all directions orthogonal to the modal axis have equal weight and so this model implicitly assumes a spherical error distribution, as would be obtained from independent isotropic landmarks with equal variances. In this case the complex Bingham distribution can be re-parameterized so that

$$\boldsymbol{A} = \kappa\,(\boldsymbol{I} - \boldsymbol{\mu}\,\boldsymbol{\mu}^*),$$

where $\boldsymbol{\mu}$ is the modal vector on the pre-shape sphere. This equivalence is used in the random generation process of the $\mathbb{C}W$ model by means of the $\mathbb{C}B$ distribution generation. For this reason, in the Chapter 3, we are interested at studying triangles in the plan and $\boldsymbol{A}$ is a matrix with order two.

The Proposition 2.2 suggests a parametric relationship between the $\mathbb{C}B$ and $\mathbb{C}W$ models; in particular, the $\mathbb{C}W$ concentration parameter κ is written in terms of the $\mathbb{C}B$ of vector eigenvalues.

Proposition 2.2. *Let* $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$ *(for* $k+1=3$*). Then* $\boldsymbol{A} = \kappa(\boldsymbol{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ *implies in*

$$\kappa = \frac{2\lambda_1}{g(\boldsymbol{\mu})}, \tag{2.7}$$

where $g(\boldsymbol{\mu})$ *is the function of the modal vector* $\boldsymbol{\mu}$ *given in Appendix A.*

The proof of this result can be found in Appendix A.

2.2.4 Simulation for the $\mathbb{C}B$ distribution

By reproducibility issues, we mention $\mathbb{C}W$ distributed data are generated from the $\mathbb{C}B$ model. The methods for generation algorithms on $\mathbb{C}B$ were proposed by Kent, Constable and Er (2004), which showed this process depends only from the TME distribution. In this chapter, we use the truncation to a simplex method combined with the TME model. First we simulate k outcomes from the truncated exponential, Texp(λ), by the acceptance-rejection method and then these values are expressed in polar coordinates to obtain observations of the $\mathbb{C}B$ model. The algorithm 1 represents steps to simulate the TME distribution by means of the acceptance-rejection procedure.

The method for simulating the $\mathbb{C}B$ distribution uses observations of the TME model, as follows.

The algorithm 2 returns a vector k, $\boldsymbol{z} = [z_1, \ldots, z_k]^\top$ which has a $\mathbb{C}B$ distribution.

Algorithm 1 Simulation of the truncated exponential distribution

1: Generate observations, say u_j, of a random sample from $u_j \sim U[0,1]$, $j = 1, \ldots, k-1$.
2: Let $s_j^{'} = -(1/\lambda_j)\log(1 - u_j(1 - \exp(-\lambda_j))$ and $s^{'} = [s_1^{'}, \ldots, s_{k-1}^{'}]^{\top}$, such that $s_j^{'}, j = 1, \ldots, k-1$ are outcomes of a random sample from Texp(λ_j.)
3: If $\sum_{j=1}^{k-1} s_j^{'} < 1$, set $s = s^{'}$. Otherwise reject $s^{'}$ and return to 1.

Algorithm 2 Simulation of complex Bingham distribution

1: Generate $s^{'} = [s_1^{'}, \ldots, s_{k-1}^{'}]^{\top}$, outcomes from $s_j \sim \text{Texp}(\lambda_j)$ according to the Algorithm 1.
2: If $\sum_{j=1}^{k} s_j^{'} < 1$, define $s_k = 1 - \sum_{j=1}^{k-1} s_j^{'}$. Otherwise, return to step 1.
3: Generate an observed sample, say θ_j, for $j = 1, \ldots, k$ from a random sample from $\boldsymbol{\theta} \sim U[0, 2\pi)$, $j = 1, \ldots, k$.
4: Compute $z_j = s_j^{1/2}\exp(\mathrm{i}\theta_j)$, $j = 1, \ldots, k$.

2.3 Statistical Information Theory

Information Theory (IT) is a branch of the Mathematics which is related to Probability and Statistics. It plays an important role in Modern Communication Theory. In a concise way, it can be said that its field of activity is related to the mathematical systematization of information transmission between communication channels. The first studies in IT aimed were performed by Hartley (1928). Shannon (1948) has extended this idea, considering – among other things – the effect of noise on communication channels. The crucial concepts of Shannon theory lie on two measures: information and entropy. Such concepts have a considerable impact in several areas, such as, Telecommunication and Statistical Inference Blatt and Hero (2007). From these concepts, a tool called relative entropy measurement (a kind of divergence) was introduced by Kullback (1978). This tendency of connecting Mathematical Statistics and IT has been coined as Statistical Information Theory (SIT). The first goal of relative entropy was to quantify the error of choosing a variable when another would be more adjustable to the information. Subsequently, to the first classes of entropies and divergences class have been invested in the statistical inference context.

Over the years, some divergence classes have been used in statistical inference for various

purposes, such as the study of their asymptotic properties and proposals of hypothesis tests. Esteban (1996) have provided entropy-and divergence-based goodness-of-fit and homogeneity tests can be applied to multinomial populations. Morales, Pardo and Pardo (2001) have proposed a divergence-based procedure to test composite hypotheses related to s populations when sample sizes are different. Eguchi and Copas (2006) have explored a connection between the Kullback-Leibler divergence and the Neyman–Pearson lemma. Nascimento, Cintra and Frery (2010) have derived and compared eight stochastic distances in the SAR image processing context. Taneja (2013) have furnished bounds on the probability of error in terms of generalized symmetric divergence measures. Röver and Friede (2017) have proposed a method (in terms of the KL divergence) of constructing an easily tractable discrete mixture distribution as an approximation. Contreras-Reyes (2014) have derived an asymptotic expression for the KL divergence measure of the multivariate skew-t distribution.

According to Kullback (1978), Information Theory has its mathematical roots in the concept of disorder or entropy in thermodynamics and statistical mechanics. Rényi et al. (1961) have stated that the entropy can be interpreted not only as an uncertainty measure as also an information measure. Pardo (2005) have interpreted the entropy as diversity measure that appears in many areas. In the field Economy, the diversity quantity is defined as the presence of a great number of different types of industries in a geographical area. In Biology, this measure is understand as the number of species in a place as well as the abundance of those species. In this thesis, the concept of entropy is understood as a variability multivariate measure approached by Chen et al. (2016). Some entropy classes have been used in statistical inference for various purposes, such as the study of their asymptotic properties and proposals of hypothesis tests. Lin (1991) have introduced a new class of information-theoretic divergence measures based on the Shannon entropy. He determined the relationship of such measures with the variational distance and the probability of misclassification error. Nadarajah and Zografos (2005) have derived exact expressions for Rényi and Shannon entropies for various continuous bivariate distributions. Baratpour, Ahmadi and Arghami (2008) have explored properties of the Rényi entropy of order statistics. Frery, Cintra and Nascimento (2013) have furnished analytical expressions for the Shannon, Rényi, and restricted Tsallis entropy measurements under scaled complex

Wishart distribution. Pietrzak et al. (2016) have proposed a frequency-based framework using the Gaussian limit results for the effective number of species (the Hill numbers) with the approach of the empirical Rényi entropy and divergence. Chen et al. (2016) have provided some new insights about the behaviors of entropy as a measure of multivariate variability. Seo and Kang (2014) have derived the entropy of the generalized half-logistic distribution in terms of Type-II censored samples from a Bayesian perspective.

Song, Song and Kang (2017) have developed two methods which were based on the maximum entropy principle and the ordinary entropy method for estimating the parameters of the four-parameter exponential gamma distribution. More information on these stochastic measures can be found in Rényi et al. (1961), Kailath (1967), Diaconis and Zabell (1982), Middleton, Electrical and Engineers (1996), MacKay (2003), Pardo (2005), Liese and Vajda (2006) and Emmert-Streib and Dehmer (2009).

2.3.1 Divergence measures

In this thesis, we admit the convention that a "divergence" represents any non negative function, say $D(\boldsymbol{x}||\boldsymbol{y}) \geq 0$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{Z} \subseteq \mathbb{C}^p$ and $p \in \mathbb{Z}_+$, between two probability measures that satisfy the identity of definiteness property (BARTLE; SHERBERT, 1982, p. 328, chap. 11,); i.e., $D(\boldsymbol{x}||\boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$. When such a function is also symmetric, it is called as "distance", say

$$d(\boldsymbol{x}, \boldsymbol{y}) \geq 0 : (\text{i})\, d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}\,(\text{definiteness}) \text{ and } (\text{ii})\, d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})\,(\text{symmetry}).$$

Finally, we conceive "metric" as a distance which also satisfies the triangular inequality (DEZA; DEZA, 2009). Now we particularize this theory to the $\mathbb{C}B$ and $\mathbb{C}W$ models.

Definition 2.13. *Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be random variables defined over the same probability space, equipped with densities $f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)$. The Kullback-Leibler and Rényi divergences are given by, respectively,*

$$D_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = \int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1) \log\left[\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)}\right] w_k(d\boldsymbol{z}) = \mathbb{E}_{\boldsymbol{\theta}_1}\left[\log\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)}\right)\right]$$

and

$$D_R^{\beta}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) \;=\; \frac{1}{\beta-1}\,\log\int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_1}^{\beta}(\boldsymbol{z};\boldsymbol{\theta}_1)\, f_{\boldsymbol{x}_2}^{1-\beta}(\boldsymbol{z};\boldsymbol{\theta}_2)\, w_k(d\boldsymbol{z}) \\ = \frac{1}{\beta-1}\log \mathbb{E}_{\boldsymbol{\theta}_1}\left[\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{\theta}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{\theta}_2)}\right)^{\beta-1}\right],$$

where $\beta \in (0,1)$ is the Rényi divergence order, $\mathbb{E}_{\boldsymbol{\theta}_1}(\cdot)$ is the expected value in terms of $\boldsymbol{\theta}_1$ and $w_k(d\boldsymbol{z})$ is the uniform measure on $\mathbb{C}\boldsymbol{S}^{k-1}$. Since the Kullback-Leibler divergence is a non symmetric measure, Seghouane and Amari (2007) have proposed the symmetrization of this divergence:

$$d_{KL}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \;=\; \frac{[\,D_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) \;+\; D_{KL}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)\,]}{2}.$$

We denote $d_{KL}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ as the KL distance. The quantity $2\,d_{KL}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is also known as J-divergence or Jeffreys divergence. Some strategies have been used to symmetrize modifications of $D_R^{\beta}(\cdot||\cdot)$, such as for the directed divergence proposed by Chung et al. (1989). Here, we assume

$$d_R^{\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \;=\; \frac{1}{\beta-1}\log\left\{\frac{\exp[(\beta-1)\,D_R^{\beta}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)] \;+\; \exp[(\beta-1)D_R^{\beta}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)]}{2}\right\}.$$

It is important to emphasize that throughout the thesis the value of the **order parameter** β **is not estimated but chosen**. In what follows, we define other used quantities.

Definition 2.14. *Under the same conditions in the Definition 2.13, the Bhattacharyya and Hellinger distances between the probability measures are given by, respectively,*

$$d_B(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \;=\; -\log\int_{\mathbb{C}\boldsymbol{S}^{k-1}} \sqrt{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{\theta}_1)\, f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{\theta}_2)}\, w_k(d\boldsymbol{z}) = -\log \mathbb{E}_{\boldsymbol{\theta}_1}\left[\sqrt{\frac{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{\theta}_2)}{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{\theta}_1)}}\right]$$

and

$$\tilde{d}_H(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \;=\; \sqrt{\frac{1}{2}\int_{\mathbb{C}\boldsymbol{S}^{k-1}} \left(\sqrt{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{\theta}_1)} \;-\; \sqrt{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{\theta}_2)}\right)^2 w_k(d\boldsymbol{z})} \\ = \sqrt{1-\mathbb{E}_{\boldsymbol{\theta}_1}\left[\sqrt{\frac{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{\theta}_2)}{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{\theta}_1)}}\right]}.$$

In this work, by issues of formalism with the Hellinger statistic that we propose next, we refer the Hellinger distance as $d_H(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = [\tilde{d}_H(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)]^2$.

2.3.2 Entropy measures

We discuss the definitions and some properties of these entropy measures a follows.

Definition 2.15. *Let $\boldsymbol{x}$ be a random variable having an absolutely continuous distribution with its pdf given by $f(\boldsymbol{x};\boldsymbol{\theta})$. The Shannon entropy denoted by (H_S) is defined as*

$$H_S(\boldsymbol{x}) \equiv H_S(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[-\log(f(\boldsymbol{x};\boldsymbol{\theta}))] = -\int [\log(f(\boldsymbol{x};\boldsymbol{\theta}))] f(\boldsymbol{x};\boldsymbol{\theta}) w_k(d\boldsymbol{x}).$$

The Rényi entropy with order $\beta \in \mathbb{R}_+ - \{1\}$ and denoted by (H_R^β) is expressed by

$$H_R^\beta(\boldsymbol{x}) \equiv H_R^\beta(\boldsymbol{\theta}) = \frac{1}{1-\beta} \log\left[\mathbb{E}_{\boldsymbol{\theta}}(f(\boldsymbol{x};\boldsymbol{\theta})^{\beta-1})\right] = \frac{1}{1-\beta} \log \int f(\boldsymbol{x};\boldsymbol{\theta})^\beta w_k(d\boldsymbol{x}).$$

Golshani and Pasha (2010) and Contreras-Reyes (2015) have provided some relevant properties of the Rényi entropy, listed below:

1. The Rényi entropy can be negative;

2. The Rényi entropy is invariant under a location transformation of the random variable. For example, for $m \in \mathbb{R}$

$$H_R^\beta(\boldsymbol{x} + m) = H_R^\beta(\boldsymbol{x});$$

3. The Rényi entropy is not invariant under a scale transformation of the random variable. For example, for $m \in \mathbb{R}$ we have:

$$H_R^\beta(m\boldsymbol{x}) = H_R^\beta(\boldsymbol{x}) + |\log(m)|;$$

4. For $\beta_1 < \beta_2$, we have $H_R^{\beta_1}(\boldsymbol{x}) \geq H_R^{\beta_2}(\boldsymbol{x})$ for all $\boldsymbol{x}$, and the equality holds if and only if $\boldsymbol{x}$ is a uniform random variable;

5. The Rényi entropy converge for Shannon entropy for all $\beta \in \mathbb{R}_+ - \{1\}$, that is $\lim_{\beta \to 1} H_R^\beta(\boldsymbol{x}) = H_S(\boldsymbol{x})$.

Additional properties about Rényi and Shannon entropies can be found in Pardo (2005), Cover and Thomas (1991) and Sánchez-Moreno, Angulo and Dehesa (2014).

3 DISTANCE-BASED HYPOTHESIS TESTS FOR TRIANGLE SHAPES

3.1 Introduction

This chapter focuses on SSA for pre-shape data obtained from triangles in the plan. In this context, Kendall has shown how variables for pre-shape of labeled triangles can be mapped to points on a sphere Mardia and Dryden (1989a). In terms of visualization, Kendall (1983) has defined one of the half-lunes of the shape sphere like projection area of unlabeled triangle shapes. Some papers have presented studies about models, inference and other statistical methods for pre-shapes of triangles as follows. A statistical description of mean biological form and size-changes by means of linearized significance tests has been proposed in Kendall (1984). The tests have been based on a convenient parametric distribution for shape changes in triangles. Mardia and Dryden (1989a) have presented a distribution for pre-shapes of triangles under the Bookstein model using the Mardia and Dryden method. Furthermore, it was examined the reduction of this distribution when the shape space is reduced to a subset of the sphere. Dryden and Mardia (1991) have investigated the exact distribution for pre-shapes of general Gaussian labeled point configurations in two dimensions and the triangle case has been considered in detail. Mardia (1999) has provided an uniformity test for highly dispersed shapes, using the standard techniques of directional statistics. They have approached the isometric transformation from triangular shapes to a sphere in three dimensions in order to provide a rich class of distributions for pre-shapes. Rohlf (1999) has demonstrated that the pre-shape space of planar triangles Procrustes aligned to a reference triangle corresponds to a unit hemisphere.

A common assumption which is made when formulating hypothesis tests for mean shape in the unit sphere is “samples should be strongly concentrated”. This fact is geometrically justified, since models under this assumption are asymptotically equivalent to the ones in the tangent plan (commonly more tractable). However, sometime we find data having low concentration. This chapter addresses a way by means of SIT to treat with shape change on both high and low concentrations. We propose distance-based two-sample hypothesis

tests for pre-shapes which are extracted from triangles. To obtain novel associated pivotal statistics, we derive the Hellinger (H), Bhattacharyya (B), Rényi (R) and KL distances for the $\mathbb{C}W$ model. We also prove the KL distance is invariant under rotation, but the other three outperform this drawback. We carry out Monte Carlo experiments to quantify the performance of our proposals, adopting as comparison criteria empirical test sizes and powers. Finally, an application to real data is made using the second thoracic vertebrae T2 of mice in order to assess possible effects of body weight on the mouse vertebra shape. Numerical evidences indicate that our proposals may be better detectors of mean shape difference between samples of triangles than the F2 test (analysis-of-variance tailored to SSA literature) proposed by Mardia and Dryden (1999).

The chapter is organized as follows. Section 3.2 tackles the two-sample hypothesis tests based on distance measures for pre-shape data. Numerical results are displayed in Section 3.4. The conclusion remarks are presented in Section 3.5. Finally, proofs of the theoretical results are given in Appendix A.

3.2 Distance-Based Measure for the $\mathbb{C}W$ Model

In this section, we present new contributions in distance-based statistical inference in the pre-shape context. We develop closed-form expressions for the Kullback-Leibler, Rényi, Bhattacharyya and Hellinger stochastic distances between two probability measures which come from the $\mathbb{C}W$ distribution which are enunciated in the following results.

Theorem 3.1. *Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_1, \kappa_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_2, \kappa_2)$ with pdfs given, respectively, by $f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)$. Then the Kullback-Leibler distance (denoted by d_{KL}) is expressed as*

$$d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{1}{2}\left\{ \frac{\kappa_1}{c_1(\kappa_1)} \sum_{j=1}^{k} \lambda_{T_{1j}} \frac{\partial}{\partial \lambda_{M_{1j}}} c(\boldsymbol{\Lambda}_{M_1}) - \frac{\kappa_2}{c_1(\kappa_1)} \sum_{j=1}^{k} \lambda_{M_{2j}} \frac{\partial}{\partial \lambda_{M_{1j}}} c(\boldsymbol{\Lambda}_{M_1}) \right.$$
$$\left. + \frac{\kappa_2}{c_1(\kappa_2)} \sum_{j=1}^{k} \lambda_{M_{2j}} \frac{\partial}{\partial \lambda_{T_{2j}}} c(\boldsymbol{\Lambda}_{T_2}) - \frac{\kappa_1}{c_1(\kappa_2)} \sum_{j=1}^{k} \lambda_{T_{1j}} \frac{\partial}{\partial \lambda_{T_{2j}}} c(\boldsymbol{\Lambda}_{T_2}) \right\}, \tag{3.1}$$

where $c_1(\cdot)$ and $c(\cdot)$ are the $\mathbb{C}W$ and $\mathbb{C}B$ integration constants, respectively, $\dfrac{\partial c(\cdot)}{\partial \lambda_j}$ is the first derivative Amaral, Floréz and Cysneiros (2013) of normalization constant of the $\mathbb{C}B$

model expressed as

$$\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_i} = 2\pi^k \left[\sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{a_j \exp(\lambda_j)}{(\lambda_j - \lambda_i)} + a_i \exp(\lambda_i) + a_i b_i \exp(\lambda_i) \right],$$

with $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_k)$, $a_j^{-1} = \prod_{\substack{l=1 \\ l\neq j}}^{k} (\lambda_j - \lambda_l)$ *and* $b_i = \sum_{l\neq i} \frac{1}{(\lambda_i - \lambda_l)}$.

Theorem 3.2. *Under the same conditions of the Theorem 3.1. The Rényi distance with order parameter* $\beta \in (0,1)$ *(denoted by* d_R^β*) is*

$$d_R^\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{1}{(\beta-1)} \log \left\{ \frac{1}{2} \left[\left(\frac{c_1^{\beta-1}(\kappa_2) c(\boldsymbol{V}_1)}{c_1^\beta(\kappa_1)} \right) + \left(\frac{c_1^{\beta-1}(\kappa_1) c(\boldsymbol{V}_2)}{c_1^\beta(\kappa_2)} \right) \right] \right\}, \tag{3.2}$$

where $\boldsymbol{V}_1 = (\kappa_1 \beta \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^* - \kappa_2(\beta-1)\boldsymbol{\mu}_2 \boldsymbol{\mu}_2^*)$ *and* $\boldsymbol{V}_2 = (\kappa_2 \beta \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^* - \kappa_1(\beta-1)\boldsymbol{\mu}_1 \boldsymbol{\mu}_1^*)$ *are both hermitian matrices with order* k.

Corollary 3.1. *Under the same conditions of the Theorem 3.1. The Bhattacharyya (denoted by* d_B*) and Hellinger (denoted by* d_H*) distances are given by*

$$d_B(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = -\log \left(\frac{c(\boldsymbol{A})}{\sqrt{c_1(\kappa_1) c_1(\kappa_2)}} \right) \tag{3.3}$$

and

$$d_H(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 1 - \frac{c(\boldsymbol{A})}{\sqrt{(c_1(\kappa_1) c_1(\kappa_2))}}, \tag{3.4}$$

where $\boldsymbol{A} = \left(\frac{\kappa_1}{2} \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^* + \frac{\kappa_2}{2} \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^* \right)$.

The proof of the results above can be found in Appendix A.

Figure 5 illustrates the distances obtained in the results 3.1, 3.2 and 3.1. The graph was obtained by considering the following configuration: $\boldsymbol{\lambda} = (8, 0)^\top$ (vector of eigenvalues), $k+1 = 3$ (number of landmarks) and $n = 50$ (the size of sample which is used to estimate κ and $\boldsymbol{\mu}$). From $\boldsymbol{\lambda}$ fixed, estimates for $\boldsymbol{\mu}$ and κ, say $\widehat{\boldsymbol{\mu}}$ and $\widehat{\kappa}$, are obtained from $\mathbb{C}B$ distributed generated data. We then fix $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \widehat{\boldsymbol{\mu}}$ and take $(\kappa_1, \kappa_2) = (\widehat{\kappa}, (1+\epsilon)\widehat{\kappa})$ with $\epsilon \in (-0.01, 0.01)$. We can see the Hellinger and Bhattacharyya distances are closed and, when the parameter β increases to one, the Rényi distance approaches of the KL distance. Similar behavior has been obtained in works Pardo (2005), Frery, Nascimento and Cintra (2014) and Nascimento, Cintra and Frery (2010).

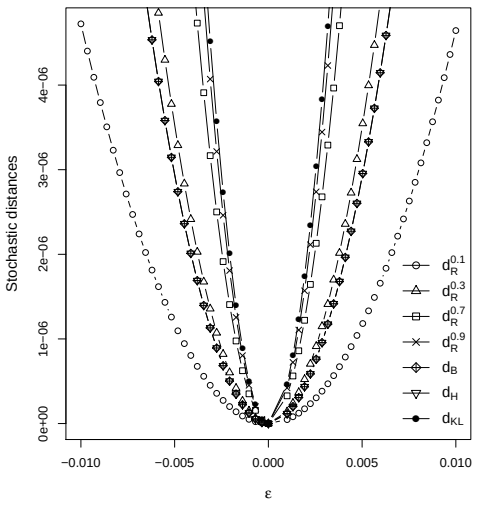


Figure 5 – Plot of KL, Rényi, Bhattacharyya and Hellinger distance measures. For graphical procedure we take $\boldsymbol{\lambda} = (8, 0)^\top$, $n = 50$ and $k + 1 = 3$. We fixed the $\boldsymbol{\mu}$ and taking $(\kappa_1, \kappa_2) = (\widehat{\kappa}, (1 + \epsilon)\widehat{\kappa})$ with $\epsilon \in (-0.01, 0.01)$

In particular, for the $\mathbb{C}W$ model, the KL distance is invariant under rotation, as discussed in the next proposition. This fact poses a disadvantage to discriminate points which are meaningfully distinct in rotation. Before we enunciate this result, a procedure for obtaining the rotation matrix described by Amaral, Dryden and Wood (2007) is discussed. Suppose that $\boldsymbol{a}$ and $\boldsymbol{b}$ are complex unit vectors in $\mathbb{C}^2$. For the application to shape analysis, a pre-shape $\tilde{\boldsymbol{b}}$ should be chosen from the shape $[\boldsymbol{b}]$ of $\boldsymbol{b}$, where $\boldsymbol{b}^*\boldsymbol{b} = 1$. Then $\tilde{\boldsymbol{b}}$ moves to $\boldsymbol{a}$ along a horizontal geodesic in the pre-shape space. Let $\tilde{\boldsymbol{b}} = \boldsymbol{b}(\boldsymbol{b}^*\boldsymbol{a})/|\boldsymbol{b}^*\boldsymbol{a}|$ so that $\tilde{\boldsymbol{b}}$ has the same norm as $\boldsymbol{b}$ and $\tilde{\boldsymbol{b}}^*\boldsymbol{a}$ is real. Now, define the vector

$$\tilde{\boldsymbol{c}} = \frac{\tilde{\boldsymbol{b}} - \boldsymbol{a}(\boldsymbol{a}^*\tilde{\boldsymbol{b}})}{|\tilde{\boldsymbol{b}} - \boldsymbol{a}(\boldsymbol{a}^*\tilde{\boldsymbol{b}})|},$$

the matrix $\tilde{\boldsymbol{A}} = \boldsymbol{a}\tilde{\boldsymbol{c}}^* - \tilde{\boldsymbol{c}}\boldsymbol{a}^*$ and the angle $\tilde{\zeta} = \cos^{-1}(\boldsymbol{a}^*\tilde{\boldsymbol{b}})$. Consider that the vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^2$ satisfy $|\boldsymbol{a}|^2 = 1$ and $|\boldsymbol{b}|^2 = 1$, and suppose $|\boldsymbol{b}^*\boldsymbol{a}| < 1$. Let $\tilde{\boldsymbol{c}}$, $\tilde{\boldsymbol{A}}$ and $\tilde{\zeta}$ be defined as above. Then the matrix $\boldsymbol{R} = \exp(\tilde{\zeta}\tilde{\boldsymbol{A}})$ with

$$\exp(\tilde{\zeta}\tilde{\boldsymbol{A}}) \equiv \boldsymbol{I}_2 + \sum_{j=1}^{\infty} \frac{\tilde{\zeta}^j}{j!}\tilde{\boldsymbol{A}}^j$$

satisfies the following properties:

(a) $\boldsymbol{R}$ is 2×2 unitary matrix; (b) $\boldsymbol{R}$ can be written

$$\boldsymbol{R} = \boldsymbol{I}_2 + \sin(\tilde{\zeta})\tilde{\boldsymbol{A}} + (\cos(\tilde{\zeta}) - 1)(\boldsymbol{a}\boldsymbol{a}^* + \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^*);$$

(c) $\boldsymbol{R}\tilde{\boldsymbol{b}} = \boldsymbol{a}$; (d) For any $\boldsymbol{z} \in \mathbb{C}^2$ such that $\boldsymbol{a}^*\boldsymbol{z} = 0 = \boldsymbol{b}^*\boldsymbol{z}$, the matrix $\boldsymbol{R}$ is the identity transformation, i.e., $\boldsymbol{R}\boldsymbol{z} = \boldsymbol{z}$.

Proposition 3.1. *Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_1, \kappa_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_2, \kappa_2)$. Let $\boldsymbol{R}$ be a rotation matrix, then the Kullback-Leibler distance is invariant under rotation; that is,*

$$d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}).$$

The proof is given in Appendix A.

3.2.1 New distance-based hypothesis tests

In what follows, we define the mathematical mechanism of the use of the (h, φ)-divergence class, on which the new hypothesis tests are proposed. A way to make hypothesis tests is through the (h, φ)-divergence class, which has been introduced by Menéndez et al. (1995) and discussed by Salicrú et al. (1994). It is an extension of the φ-divergence class – pioneered by Csiszár (1967) – and defined as follows.

Definition 3.1. *Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_1, \kappa_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_2, \kappa_2)$ be two probability measures having densities $f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)$, respectively. The (h, φ)-divergence, say $d^h_\varphi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, is defined as*

$$d^h_\varphi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h\left(\int_{\mathbb{C}S^{k-1}} f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)\, \varphi\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)}\right) w_k(d\boldsymbol{z})\right) = h\left(\mathbb{E}_{\boldsymbol{\theta}_2}\left[\varphi\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{\theta}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{\theta}_2)}\right)\right]\right), \tag{3.5}$$

where $h\colon (0, \infty) \to [0, \infty)$ is a strictly increasing (or decreasing) function with $h(0) = 0$, and $\varphi\colon (0, \infty) \to [0, \infty)$ is a convex (or concave) function such that $\varphi(1) = 0$, $0\,\varphi(0/0) = 0$ and $0\,\varphi(x/0) = \lim_{x\to\infty} \varphi(x)/x$ (SALICRU et al., 1993)

Notice that each pair (h, φ) in (3.5) yields a particular divergence. Table 2 displays those we use in this chapter. The measures (3.1), (3.2), (3.3) and (3.4) can be used isolated from any inferential context (such as in many machine learning and computer science works), but here we are also interested to study the their asymptotic behavior.

Now we are in position of proposing distance-based two-sample hypothesis tests for difference mean shapes. We are interested in to test:

$$H_0 : [\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2] \quad vs. \quad H_1 : [\boldsymbol{\mu}_1] \neq [\boldsymbol{\mu}_2], \tag{3.6}$$

Table 2 – (h, φ)-distances and their functions

(h, φ)-distance	$h(y)$	$\varphi(x)$
Kullback-Leibler	$y/2$	$(x-1)\log x$
Rényi (order β)	$\frac{1}{\beta-1}\log(\frac{(\beta-1)}{2}y+1)$	$\frac{x^\beta + x^{1-\beta} - x - 1}{\beta - 1}$
Bhattacharyya	$-\log(-y+1)$	$-\sqrt{x} + \frac{x+1}{2}$
Hellinger	$y/2$	$(\sqrt{x}-1)^2$

where $[\boldsymbol{\mu}_j] = \{\exp(\mathrm{i}\theta)\boldsymbol{\mu} : 0 \leq \theta < 2\pi\}$, for $j = 1, 2$ (i.e. $[\boldsymbol{\mu}_j]$ represents the shape corresponding to the modal pre-shape $\boldsymbol{\mu}_j$). For obtaining of the novel pivotal statistics, under the regularity conditions discussed and proved in Salicrú et al. (1994), we use the following lemma holds.

Lemma 3.1. *Let* $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\theta}_{11}, \ldots, \widehat{\theta}_{1M})$ *and* $\widehat{\boldsymbol{\theta}}_2 = (\widehat{\theta}_{21}, \ldots, \widehat{\theta}_{2M})$ *be ML estimators for parameter vectors* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$ *having* $M_0 < M$ *common components based on independent samples of sizes* n_1 *and* n_2*, respectively. If* $\dfrac{n_1}{n_1+n_2} \longrightarrow \delta \in (0,1)$ *when* $n_1, n_2 \to \infty$ *and* $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$*, then*

$$S_\varphi^h(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{2n_1n_2}{n_1+n_2} \frac{d_\varphi^h(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)}{\int_{\mathcal{X}} h'(0)\varphi''(1)\,\mathrm{d}\upsilon}$$

In particular for this manuscript,

$$S_D(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \nu \frac{2\,n_1\,n_2}{n_1\,+\,n_2}\, d_D(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_{M-M_0},$$

where υ *is the* σ*-finite measure on the measurable space* $(\mathcal{X}, \mathcal{B})$*, "*$\xrightarrow[n_1,n_2\to\infty]{\mathcal{D}}$*" represents the* convergence in distribution*,* $D = \{KL, R, B, H\}$ *and* ν *is a chosen value according to the used stochastic distance; i.e.,* $\nu = [h'(0)\varphi''(1)]^{-1} = \{1, 1/\beta, 4, 4\}$*. In what follows, we assume* $\mathcal{X} = \{1\}$ *and* $\upsilon(1) = 1$*, which yields*

$$\int_{\mathcal{X}} h'(0)\varphi''(1)\,\mathrm{d}\upsilon = h'(0)\varphi''(1)$$

(Salicrú et al., 1994, assumption adopted in Example 1).

Based on Lemma (4.1), tests for the null hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ can be derived, as given in the next proposition.

Proposition 3.2. *Let* n_1 *and* n_2 *be large and* $S_D(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = s$*, then the null hypothesis* $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$*, can be rejected at level* α *if* $P(s < \chi^2_{M-M_0}) \leq \alpha$*.*

From Lemma 4.1 and Proposition 4.2 combined with the Theorem 5.1, four new hypothesis tests for the $\mathbb{C}W$ distribution come from distance-based pivotal statistics which we provide subsequently. We divide this discussion in two cases:

Case 1 ($\kappa_1 = \kappa_2$):

[1.1] Rényi statistics with order $\beta \in (0,1)$ - $S_R^\beta(\cdot,\cdot)$:

$$S_R^\beta(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \beta^{-1}\frac{2n_1n_2}{n_1+n_2} d_R^\beta(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_2. \tag{3.7}$$

[1.2] Bhattacharyya statistics - $S_B(\cdot,\cdot)$:

$$S_B(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{8n_1n_2}{n_1+n_2} d_B(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_2.$$

[1.3] Hellinger statistics - $S_H(\cdot,\cdot)$:

$$S_H(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{8n_1n_2}{n_1+n_2} d_H(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_2.$$

From adopting $M_0 = 1$ in Proposition 4.2, the S_R^β, S_B and S_H statistics in the case 1 follow asymptotically the chi-squared distribution with degrees of freedom two. This discussion does not include the KL distance because its associated distance is invariant under rotation, as proven in Proposition 3.1.

Case 2 ($\kappa_1 \neq \kappa_2$):

[2.1] Kullback-Leibler statistics - $S_{KL}(\cdot,\cdot)$:

$$S_{KL}(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{2n_1n_2}{n_1+n_2} d_{KL}(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_1.$$

[2.2] Rényi statistics with order $\beta \in (0,1)$ - $S_R^\beta(\cdot,\cdot)$:

$$S_R^\beta(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \beta^{-1}\frac{2n_1n_2}{n_1+n_2} d_R^\beta(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_3.$$

[2.3] Bhattacharyya statistics - $S_B(\cdot,\cdot)$:

$$S_B(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{8n_1n_2}{n_1+n_2} d_B(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_3.$$

[2.4] Hellinger statistics - $S_H(\cdot,\cdot)$:

$$S_H(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{8n_1n_2}{n_1+n_2} d_H(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_3. \tag{3.8}$$

In the case 2, the S_{KL} statistic follows the χ^2_1 model because - after matrix manipulations involving quadratic forms - its distance in (3.1) is only the result of comparison between two non null eigenvalues. The rest of them have the χ^2_3 model from plunging $M_0 = 0$ in Proposition 4.2.

3.3 Asymptotic Results under High Concentration

Mardia and Dryden (1999) have proposed theoretical and asymptotic results for the $\mathbb{C}W$ probability measure under large concentration. Now we highlight some of these results that we are going to use in section of numerical results. The next proposition defines a limit relation between $\mathbb{C}W$ and complex normal distributions.

Proposition 3.3. *If κ is large, then the $\mathbb{C}W$ model tends to the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $(1/2\kappa)(\boldsymbol{I}_k - \boldsymbol{\mu}\boldsymbol{\mu}^*)^-$, where $(\cdot)^-$ represents the generalized inverse.*

From Proposition 3.3,

$$2\kappa[(\boldsymbol{z}-\boldsymbol{\mu})^*(\boldsymbol{I}_k - \boldsymbol{\mu}\boldsymbol{\mu}^*)(\boldsymbol{z}-\boldsymbol{\mu})] = 2\kappa(1 - \boldsymbol{z}^*\boldsymbol{\mu}\boldsymbol{\mu}^*\boldsymbol{z}) = 2\kappa \sin^2(\rho(\boldsymbol{H}^\top \boldsymbol{z}, \boldsymbol{H}^\top \boldsymbol{\mu}))$$

follows approximately the χ^2_{2k-4} distribution for large κ, since the complex rank $(\boldsymbol{I}_k - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ is $k-2$. Then, it holds

$$S_1 = 2\kappa \sin^2(\rho(\boldsymbol{H}^\top \boldsymbol{z}, \boldsymbol{H}^\top \boldsymbol{\mu})) \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \chi^2_{2k-4}. \tag{3.9}$$

In the next section, we will use (3.9) as a criterion to identify points which indicate low and high concentration degrees within the $\mathbb{C}W$ parametric space.

Mardia and Dryden (1999) have developed a large concentration two-sample test to detect different between shapes using the Equation (3.9) as follows. Let $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n$ and $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m$ be two independent random samples from $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_1, \kappa_1)$ and $\boldsymbol{y} \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}_2, \kappa_2)$, respectively.

Suppose that one wants to test the hypothesis (4.3). Using the equation (3.9) for large κ, it follows that

$$\sum_{i=1}^{n} \sin^2[\rho(\boldsymbol{z}_i, \widehat{\boldsymbol{\mu}}_1)] + \sum_{j=1}^{m} \sin^2[\rho(\boldsymbol{y}_j, \widehat{\boldsymbol{\mu}}_2)] \xrightarrow[n_1,n_2\to\infty]{\mathcal{D}} \frac{1}{2\kappa}\chi^2_{(2k-4)(n+m-2)}.$$

By analogy with analysis-of-variance, we obtain

$$\sum_{i=1}^{n}\sin^2[\rho(\boldsymbol{z}_i,\widehat{\widehat{\boldsymbol{\mu}}}_1)]+\sum_{j=1}^{m}\sin^2[\rho(\boldsymbol{y}_j,\widehat{\widehat{\boldsymbol{\mu}}}_1)]\xrightarrow[n_1,n_2\to\infty]{\mathcal{D}}\frac{1}{2\kappa}\chi^2_{(2k-4)(n+m-1)},$$

where $\widehat{\widehat{\boldsymbol{\mu}}}_1$ is the overall ML estimator for $\boldsymbol{\mu}_1$ if the two groups are pooled.

Setting

$$B=\sum_{i=1}^{n}\sin^2[\rho(\boldsymbol{z}_i,\widehat{\widehat{\boldsymbol{\mu}}}_1)]+\sum_{j=1}^{m}\sin^2[\rho(\boldsymbol{y}_j,\widehat{\widehat{\boldsymbol{\mu}}}_1)]-\sum_{i=1}^{n}\sin^2[\rho(\boldsymbol{z}_i,\widehat{\boldsymbol{\mu}}_1)]-\sum_{j=1}^{m}\sin^2[\rho(\boldsymbol{y}_j,\widehat{\boldsymbol{\mu}}_2)],$$

we have (under H_0)

$$S_{F2}=\frac{(n+m-2)B}{\sum_{i=1}^{n}\sin^2[\rho(\boldsymbol{z}_i,\widehat{\boldsymbol{\mu}}_1)]+\sum_{j=1}^{m}\sin^2[\rho(\boldsymbol{y}_j,\widehat{\boldsymbol{\mu}}_2)]}\xrightarrow[n_1,n_2\to\infty]{\mathcal{D}}F_{2k-4,(2k-4)(n+m-1)},\tag{3.10}$$

which may be defined as decision rule that rejects H_0 for large values of S_{F2}.

According to Mardia and Dryden (1999), by using Taylor series expansions for large concentrations,

$$B\approx(n^{-1}+m^{-1})^{-1}\sin^2[\rho(\widehat{\boldsymbol{\mu}}_1,\widehat{\boldsymbol{\mu}}_2)].$$

For large κ, the test based on S_{F2} is equivalent to the two-sample test proposed by Goodall (1991).

3.4 Numerical Results and discussion

This section addresses a numerical assessment of the performance of the proposed tests, comparatively to the S_{F2} test of the SSA literature. This study has been made over both generated $\mathbb{C}W$ and real data.

3.4.1 Analysis with synthetic data

In order to assess the new hypothesis tests which are induced from the combination between Theorem 5.1 and Proposition 4.2, we have employed $\mathbb{C}W$ generated data, using the simplex method proposed by Kent, Constable and Er (2004) and the mapping given in (2.7). An important practical procedure in SSA (e.g., for making hypothesis tests) is to understand if an under-study phenomenon is on low or high concentration. Thus, before

we carry out a detailed study, a pilot essay has been made to investigate when pre-shape data are strongly concentrated in terms of Proposition 3.3; i.e.,

$$H_0 : \mathbb{C}W \stackrel{d}{=} \text{Complex normal (concentrated strongly data)},$$

where $\stackrel{d}{=}$indicates equivalence in distribution. In particular, we can use the limit result of S_1 in (3.9). For this initial quantitative study, we also used the Kolmogorov-Smirnov (KS) test (DARLING, 1957) at the level significance 5%. Based on Monte Carlo experiments with three thousand replicas, we have furnished values of S_1 as input to the KS test. As prior knowledge, we suspected that scenarios which are governed by indexes (i) $\lambda = \{(1.5, 0), (2, 0)\}$ and (ii) $\lambda = \{(8, 0), (15, 0)\}$ represent low and high concentration degrees, respectively. Figure 6 illustrates such scenarios on the unit sphere through the plot of four generated samples. Indeed, (i) and (ii) point out by visual inspection low and high concentrations, respectively.

Note that identifying similarity before low concentration degrees is intuitively more difficult than in concentrated strongly scenarios. Table 3 displays results of mean KS

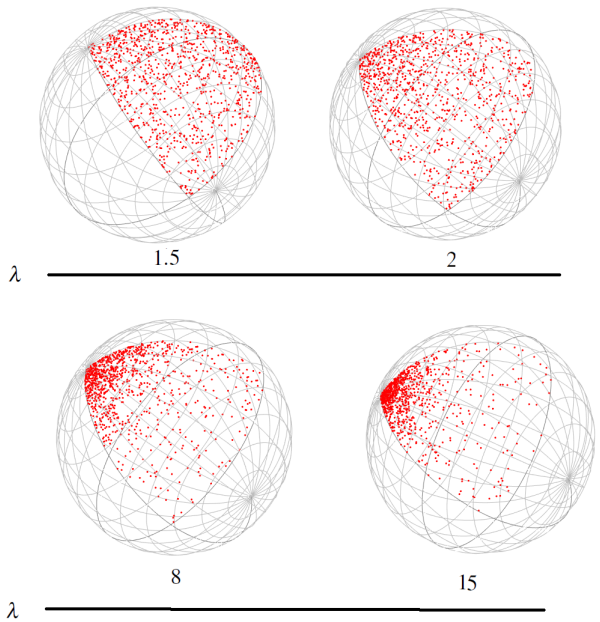


Figure 6 – Illustration of the low and high concentrations scenarios, respectively. Considering the eigenvalues vectors (i) $\lambda = \{(1.5, 0), (2, 0)\}$ and (ii) $\lambda = \{(8, 0), (15, 0)\}$.

test p-values ($\bar{\hat{\alpha}}_{KS}$) for three sample sizes ($n = 50, 70, 100$) and the situations of Figure 6. Results indicate quantitatively samples indexed at $\kappa = 1.5$ and 2 as of low concentration

and those indexed at $\kappa = 8$ and 15 as of high concentration. In the rest of this section, we assume the last conclusion.

Table 3 – Mean KS test p-value for four concentration degrees

Low concentration			High concentration		
λ	n	$\bar{\widehat{\alpha}}_{KS}$	λ	n	$\bar{\widehat{\alpha}}_{KS}$
(1.5,0)	50	0.0047	(8,0)	50	0.3453
	70	0.0033		70	0.2998
	100	5.6230×10^{-5}		100	0.2148
(2,0)	50	0.0480	(9,0)	50	0.3705
	70	0.0321		70	0.2504
	100	0.0034		100	0.2348

Now, we are in position of performing the size and power studies of the proposed two-sample tests in contrast with the S_{F_2} test of the SSA literature. We have considered the cases 1 and 2 as well as two concentration degrees, low and high. We have investigated initially the asymptotic behavior of the proposed pivotal statistics by means of a short study. As discussed, we expect the pivotal statistics S_R^β, S_B and S_H in the case 1 converge in distribution (under the null hypothesis) for χ^2_2. In numerical terms, we considered 3000 Monte Carlo replicas, on which values of S_H for two samples with $n_1 = n_2 = 70$ under the high concentration scenario $\lambda = (8, 0)$ were obtained. The KS test was employed to an array of observations of S_H, yielding p-value $\widehat{\alpha}_{KS} = 0.7162$ and concluding the observed array comes from the χ^2_2 at level 5%. The S_{F_2} test has also been used over the same data, indicating the same conclusion ($\widehat{\alpha}_{KS} = 0.6891$). In order to illustrate this behavior, Figure 7 displays the theoretical and empirical densities for S_H and S_{F2} statistics as well as their respective quantile-quantile plots (qq-plots). Graphics indicate both statistics work well.

For the case 2, we look for the pivotal statistics S_R^β, S_H and S_B (under the null hypothesis) converge in distribution for χ^2_3 and S_{KL} for χ^2_1. The KS test has been employed to outcomes of $S_R^{0.9}$ and S_H at $n_1 = n_2 = 100$ and $\lambda = (8, 0)$, yielding p-values $\widehat{\alpha}_{KS} = 0.9382$ and $\widehat{\alpha}_{KS} = 0.9973$, respectively. The same investigating mechanism applied to observations of S_{KL} has provided $\widehat{\alpha}_{KS} = 0.8322$. Figure 8 illustrates the asymptotic behavior of these statistics.

All tools reach what is expected. Subsequently, tests we discussed in previous section are submitted to wider studies.

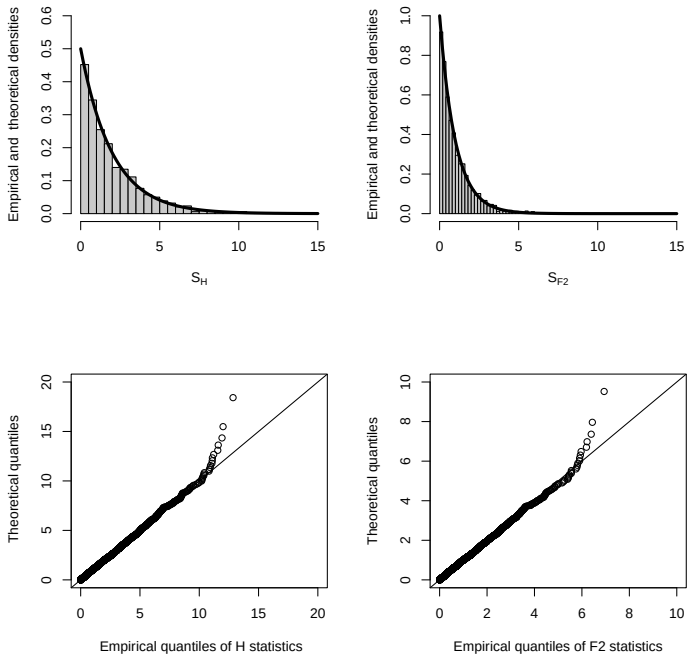


Figure 7 – Histograms and qq-plots of Hellinger and F_2 statistics for $\kappa_1 = \kappa_2$ under high concentration. The configurations considered for this simulate procedure were: the eigenvalues vector $\lambda = (8, 0)$ and sample size $n_1 = n_2 = 70$

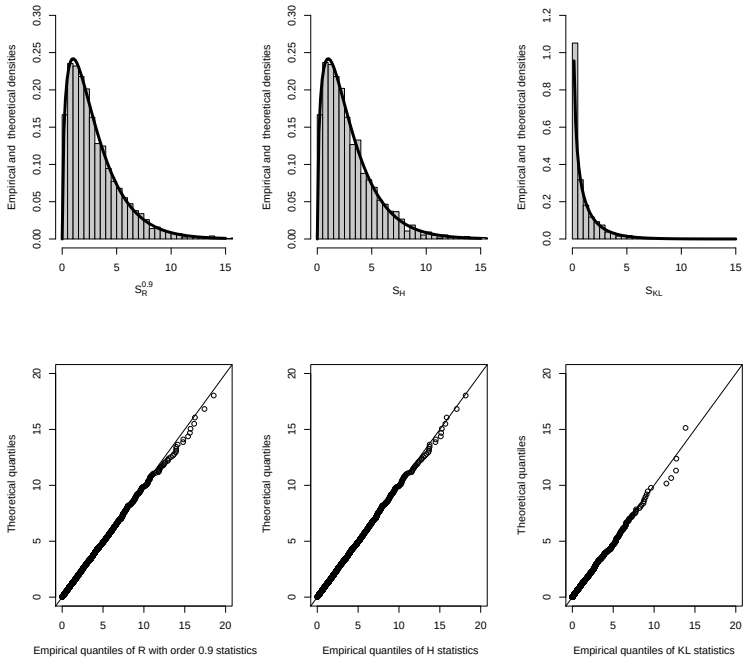


Figure 8 – Histograms of the Rényi, Hellinger and KL proposed statistics for $\kappa_1 \neq \kappa_2$ under high concentration. For graphical procedure taking: the eigenvalues vector $\lambda = (8, 0)$ and the sample size $n_1 = n_2 = 100$

Tables 4 and 5 present empirical test sizes under low concentration scenarios for the cases 1 and 2, respectively. In Table 4, rejection rates underestimate the expected value $\alpha = 5\%$. This case is the hardest situation among those considered, because it offers only the difference between mean shapes (for triangles, one eigenvalue per sample) as mechanism for decision rule in a context on which data are strongly mixtured. The lower the concentration levels, the larger sample sizes are required for tests work as expected. Although of the F2 test statistic presents a limitation in the low concentration scenario, we use this statistic in the Table 4. We can verify that the F2 test is extremely liberal for all sample sizes considered at a adopted nominal level $\alpha = 5\%$. We conclude that the application of the F2 test statistic is indicated only when the following restrictions are assumed: the concentration degrees in two samples are equal and over the high concentration scenario.

Table 5 displays the p-values of all proposed tests for case 2. Note the KL test performance was also included, since it only can be applied when the concentration parameters are assumed different. Here we verify the Hellinger and KL tests stand out for all sample sizes, showing rejection rates which are more closed to the adopted nominal level. These results are extremely important, since the majority of SSA tools assume to be under a high concentration scenario (MARDIA; DRYDEN, 1999; MICHEAS; DEY, 2005a; MICHEAS; DEY, 2005b).

Table 4 – Rejection rates under H_0 for distance-based (Hellinger, Rényi and Bhattacharyya) tests with $\kappa_1 = \kappa_2$ for low concentration scenario. The adopted nominal level $\alpha = 5\%$.

λ	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{F2}}$
(1.5,0)	30	30	0.011	0.012	0.012	0.011	0.012	0.01	0.176
	30	50	0.016	0.016	0.016	0.016	0.016	0.016	0.166
	70	70	0.026	0.027	0.027	0.026	0.027	0.026	0.183
	100	150	0.035	0.035	0.035	0.035	0.035	0.034	0.178
	200	200	0.038	0.038	0.038	0.038	0.038	0.038	0.19
(2,0)	30	30	0.022	0.022	0.022	0.022	0.022	0.019	0.127
	30	50	0.028	0.028	0.028	0.028	0.028	0.026	0.114
	70	70	0.036	0.036	0.036	0.036	0.037	0.034	0.131
	100	150	0.036	0.036	0.036	0.036	0.036	0.036	0.12
	200	200	0.04	0.04	0.04	0.04	0.04	0.04	0.132

Table 6 displays empirical test sizes for the high concentration scenario in the case 1.

Table 5 – Rejection rates under H_0 for distance-based (Hellinger, Rényi, Bhattacharyya and Kullback-Leibler) and F2 tests with $\kappa_1 \neq \kappa_2$ for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{KL}}$
(1.5,0)	30	30	0.054	0.056	0.056	0.054	0.056	0.05	0.032
	30	50	0.054	0.055	0.055	0.054	0.056	0.05	0.046
	70	70	0.05	0.052	0.052	0.05	0.052	0.048	0.043
	100	150	0.051	0.051	0.051	0.051	0.051	0.05	0.049
	200	200	0.046	0.046	0.046	0.046	0.046	0.046	0.041
(2,0)	30	30	0.054	0.055	0.055	0.054	0.055	0.049	0.039
	30	50	0.056	0.057	0.057	0.056	0.057	0.049	0.05
	70	70	0.05	0.051	0.051	0.05	0.051	0.048	0.046
	100	150	0.05	0.05	0.05	0.05	0.05	0.049	0.052
	200	200	0.044	0.044	0.044	0.044	0.044	0.044	0.044

We compare the results of the tests based on S_R^β, S_B and S_H with that on S_{F2}. We verify that they work equivalently under large values of κ. In Table 7, displays empirical test sizes for the high concentration scenario in the case 2. We can see for small sample sizes, the p-value underestimates the nominal level of the test, but for large sample sizes the tests are more significants. In summary, according to the simulation procedures performed with respect to the significance of the distance-based measures tests, for case 1 we indicate the Bhattacharyya and Rényi (order $\beta = 0.9$) statistics. For case 2 we indicate the Hellinger statistic.

Table 6 – Rejection rates under H_0 for distance-based (Hellinger, Rényi and Bhattacharyya) and F2 tests with $\kappa_1 = \kappa_2$ for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{F_2}}$
(8,0)	30	30	0.05	0.05	0.05	0.05	0.05	0.047	0.051
	30	50	0.046	0.046	0.046	0.046	0.047	0.044	0.049
	70	70	0.05	0.05	0.05	0.05	0.05	0.048	0.051
	100	150	0.043	0.043	0.043	0.043	0.043	0.042	0.045
	200	200	0.052	0.052	0.052	0.052	0.052	0.051	0.053
(9,0)	30	30	0.05	0.051	0.051	0.05	0.051	0.046	0.052
	30	50	0.047	0.048	0.048	0.047	0.048	0.044	0.049
	70	70	0.05	0.05	0.05	0.05	0.05	0.049	0.051
	100	150	0.044	0.044	0.044	0.044	0.044	0.043	0.045
	200	200	0.052	0.052	0.052	0.052	0.052	0.051	0.053

Tables 8 and 9 extend the previous study by means of empirical test power; that is, the

Table 7 – Rejection rates under H_0 for distance-based (Hellinger, Rényi, Bhattacharyya and Kullback-Leibler) tests with $\kappa_1 \neq \kappa_2$ for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{KL}}$
(8,0)	30	30	0.063	0.061	0.061	0.063	0.06	0.055	0.055
	30	50	0.059	0.058	0.058	0.059	0.058	0.052	0.062
	70	70	0.057	0.056	0.056	0.057	0.056	0.054	0.057
	100	150	0.055	0.055	0.055	0.055	0.055	0.053	0.057
	200	200	0.05	0.05	0.05	0.05	0.05	0.05	0.047
(9,0)	30	30	0.063	0.062	0.062	0.063	0.061	0.055	0.054
	30	50	0.058	0.057	0.057	0.058	0.057	0.052	0.063
	70	70	0.056	0.056	0.056	0.056	0.056	0.053	0.056
	100	150	0.056	0.056	0.056	0.056	0.056	0.054	0.056
	200	200	0.051	0.051	0.051	0.051	0.051	0.05	0.047

rejection rates when samples come from different distributions, denoted as $1 - \hat{\eta}$, where $\hat{\eta}$ is the size estimate of error of kind II. For this, it is necessary to determine a mechanism of separation of H_0 in H_1. In this thesis, we employ a procedure of transformation of the null hypothesis described by Amaral, Dryden and Wood (2007). It consists in obtaining a rotation matrix $\boldsymbol{R}$ that – when applied to the data – allows a sample to have the same mean shape, but distinct rotations. The angle used in the rotation matrix, say ϕ_{power}, assumes values in the interval $[0, 2\pi)$. We select three values of ϕ_{power}, say $0.04\pi, 0.09\pi$ and 0.3π, to illustrate the empirical test power for the case 1 in Table 8 and for the case 2 in Table 9. For $\phi_{power} \in \{0.09\pi, 0.3\pi\}$, it is noticeable tests are equivalently powerful. With exception of $(\phi_{power}, n) = (0.04\pi, 70)$, the F2 test presents rates mildly higher than others. However, we would like to highlight two advantages of distance-based tests in contrast to F2. First the statistics (3.10) depends of pre-shape outcomes, differently of ones (4.5)–(3.8). It imposes the former to higher degrees of computational complexity and of influence of potential outliers. Second distance-based tests do not have restriction of working only under high concentration.

Figure 9(a) displays the power function of the Rényi statistics with order $\beta = 0.9$ (all the others present the similar behavior) for the case 1 under the low concentration scenario. The angle vector for all graphics of power functions takes values of ϕ in $[0.001\pi, 0.005\pi, 0.009\pi, 0.01\pi, 0.02\pi, 0.03\pi, 0.04\pi, 0.05\pi, 0.06\pi, 0.07\pi, 0.08\pi, 0.09\pi, 0.1\pi, 0.2\pi, 0, 3\pi]$. Figure 9(b) tackles situations in the low concentration scenario for the case

Table 8 – Rejection rates under H_1 for distance-based (Rényi, Bhattacharyya and Hellinger) and F2 tests with $\kappa_1 = \kappa_2$ for high concentration scenario. The adopted nominal level $\alpha = 5\%$

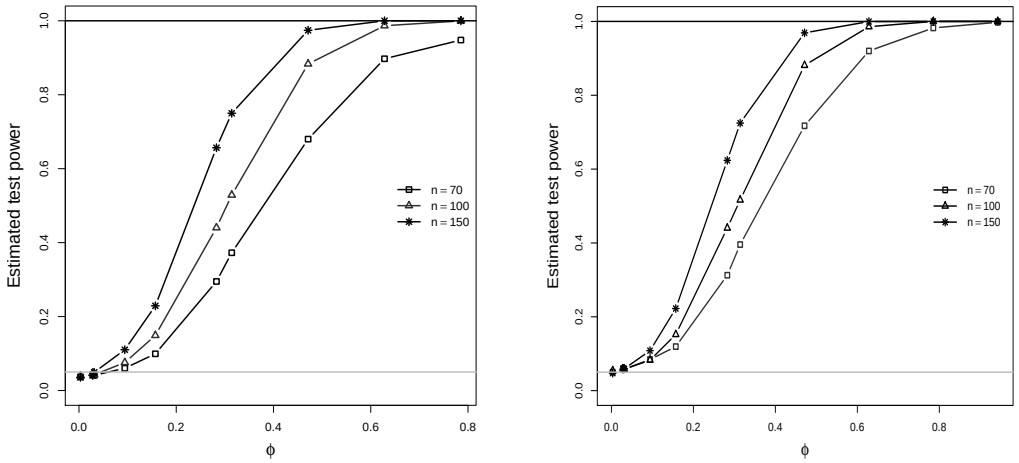
λ	ϕ_{power}	n_1	n_2	$(1-\hat{\eta})_{S_R^{0.1}}$	$(1-\hat{\eta})_{S_R^{0.3}}$	$(1-\hat{\eta})_{S_R^{0.7}}$	$(1-\hat{\eta})_{S_R^{0.9}}$	$(1-\hat{\eta})_{S_B}$	$(1-\hat{\eta})_{S_H}$	$(1-\hat{\eta})_{S_{F2}}$
(8,0)	0.04π	30	30	0.291	0.292	0.292	0.291	0.292	0.279	0.300
		70	70	0.631	0.632	0.632	0.631	0.632	0.627	0.637
		100	150	0.859	0.859	0.859	0.859	0.859	0.858	0.861
	0.09π	30	30	0.924	0.924	0.924	0.924	0.925	0.918	0.928
		70	70	0.999	0.999	0.999	0.999	0.999	0.999	0.999
		100	150	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3π	30	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	70	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		100	150	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 9 – Rejection rates under H_1 for distance-based (Hellinger, Rényi and Bhattacharyya) tests with $\kappa_1 \neq \kappa_2$ for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	ϕ_{power}	n_1	n_2	$(1-\hat{\eta})_{S_R^{0.1}}$	$(1-\hat{\eta})_{S_R^{0.3}}$	$(1-\hat{\eta})_{S_R^{0.7}}$	$(1-\hat{\eta})_{S_R^{0.9}}$	$(1-\hat{\eta})_{S_B}$	$(1-\hat{\eta})_{S_H}$
(8,0)	0.04π	30	30	0.286	0.286	0.286	0.286	0.286	0.2869
		70	70	0.592	0.591	0.591	0.592	0.591	0.583
		100	150	0.821	0.821	0.821	0.821	0.821	0.819
	0.09π	30	30	0.901	0.902	0.902	0.901	0.902	0.893
		70	70	1.00	1.00	1.00	1.00	1.00	1.00
		100	150	1.00	1.00	1.00	1.00	1.00	1.00
	0.3π	30	30	1.00	1.00	1.00	1.00	1.00	1.00
		70	70	1.00	1.00	1.00	1.00	1.00	1.00
		100	150	1.00	1.00	1.00	1.00	1.00	1.00

2. The Rényi distance test (with order $\beta = 0.9$) presents the fast increasing of the power function.

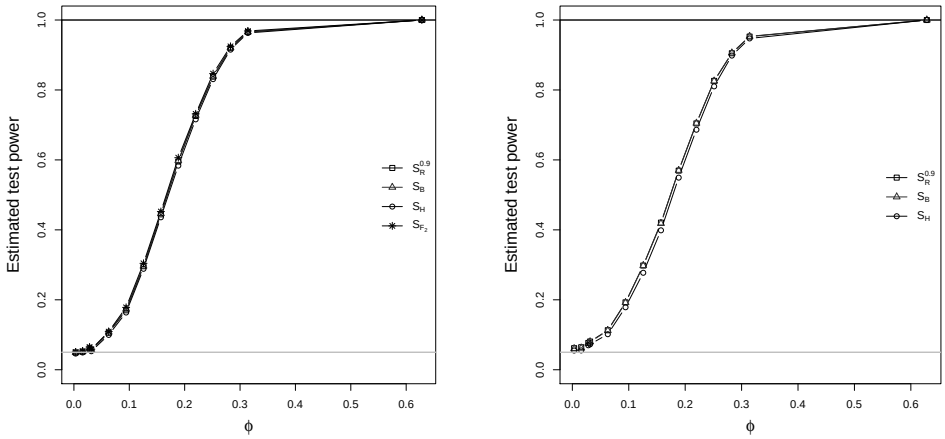
Figure 10 illustrates the empirical test power under high concentration for the cases 1 and 2, respectively. For the case 1, the proposed tests are as powerful as the F2 test. Additionally, for the case 2, we verify the good performance for all proposed tests, with exception of S_{KL} (due to its property of invariance under rotation). This means that all the proposed pivotal statistics are very useful to detect the differences between the mean shapes on the unit complex sphere for two independent samples coming from the $\mathbb{C}W$ model.



(a) Power function of the Rényi (with order $\beta = 0.9$) test for $\kappa_1 = \kappa_2$

(b) Power function of the Rényi (with order $\beta = 0.9$) test for $\kappa_1 \neq \kappa_2$

Figure 9 – Empirical power function of the Rényi (with order $\beta = 0.9$) test under low concentration for three sample sizes $n_1 = n_2 = n \in \{70, 100, 150\}$, eigenvalues vector $\lambda = (8, 0)$ and the angle selected to compute the rotation matrix ϕ



(a) Power function for distance-based and F2 tests considering $\kappa_1 = \kappa_2$

(b) Power function for distance-based tests considering $\kappa_1 \neq \kappa_2$

Figure 10 – Empirical power function of distance-based and F2 tests under high concentration for sample sizes $n_1 = 30$ and $n_2 = 20$, eigenvalues vector $\lambda = (8, 0)$ and the angle selected to compute the rotation matrix ϕ

3.4.2 Biological data analysis

Consider we wish to explore in a biology experiment the effect of the body weight on the shape of mouse vertebrae, such as described in Dryden and Mardia (2016). Three groups of mice have been considered: `large` (mice selected at each generation with large

body weights), `small` (mice selected with small body weights) and `control` (unselected mice). The real data were extracted from the package `shapes` by mean of the dataset mice available in the software `R` (for more details, see <https://cran.r-project.org/web/packages/shapes/index.html>). In this dataset, 6 landmarks have been measured on each bone and partitioned in two distinct triangles. as illustrated in Figure 11 and 13.

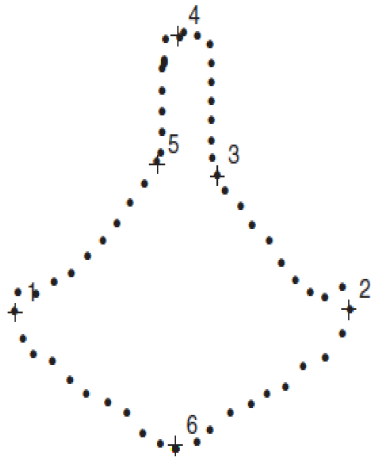


Figure 11 – Prototype of the second thoracic vertebra T2 of the group mice. Six mathematical landmarks (+) on a T2 mouse vertebra, together with 54 pseudo-landmarks around the outline. This figure was taken from Dryden and Mardia (2016, pp. 9).

The groups `small` and `large` had sample sizes of 23 bones; while the group `control`, 30 bones. This database has been utilized by many works:

1. Mardia and Dryden (1989b) have investigated shape changes in pairs of groups (`control`, `large`) and (`control`, `small`) under the exact distribution of Bookstein's variables and considering possible geometric insights.

2. Kent (1997) has compared two groups (`small` and `large`) of mouse vertebrae in order to test the difference between mean shapes using the Hotelling's T^2 test Dryden and Mardia (2016). This test demonstrated such difference is extremely significant.

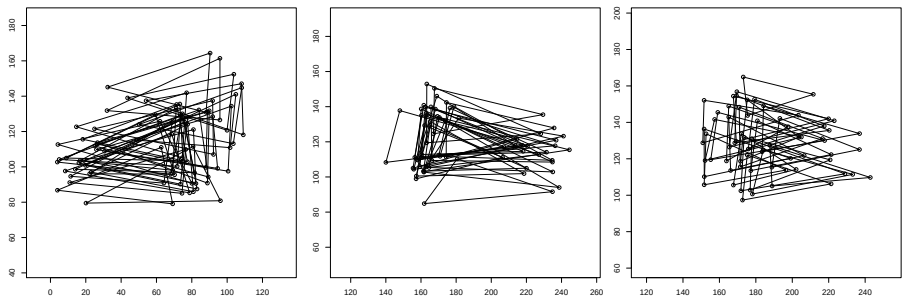
3. Micheas, Dey and Mardia (2006) have considered only the `small` group in order to make statistical inference for the mean shape by means of both maximum likelihood and Bayesian methods for the $\mathbb{C}W$ model.

More details about the biology experiment we are treated can be found in Falconer (1973), Johnson et al. (1985) and Johnson, O'Higgins and McAndrew (1988). In our application,

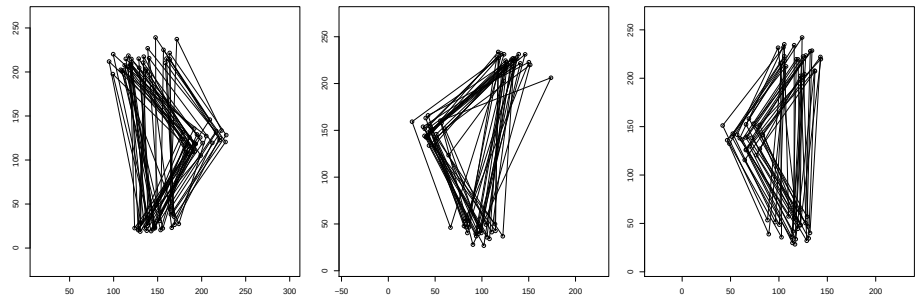
we have also considered two configurations of triangle shape extracted from the dataset of mice.

- Configuration A: triangles that are formed by landmarks 3,4 and 5 in Figure 11.
- Configuration B: triangles which are designed by landmarks 1,2 and 6 in Figure 11.

Figures 12a–12c illustrate the database of origin landmarks of the configuration A for three groups (control, large and small), and Figures 12d–12f exhibit the same for the configuration B. According to Dryden and Mardia (2016), the landmarks 1 and 2 are at maximum points of an approximate curvature function, 3 and 5 at the extreme points of negative curvature at the base of the spinous process, 4 at the tip of the spinous process, and 6 at the maximal curvature point on the opposite side of the bone from 4. The under-study objects are highly concentrated and all mutually independent Dryden and Mardia (2016).



(a) `control` in the configuration A (b) `large` in A (c) `small` in A



(d) `control` in the configuration B (e) `large` in B (f) `small` in B

Figure 12 – Databases of triangles (i.e., $k + 1 = 3$ landmarks) representing T2 vertebrae of mice at configurations A, (a)–(c), and B, (d)–(f).

In Figure 13, we use the generalized Procrustes analysis on Kent's partial tangent coordinates (DRYDEN; MARDIA, 2016) and we obtain the mean shape of the three groups for configurations A and B.

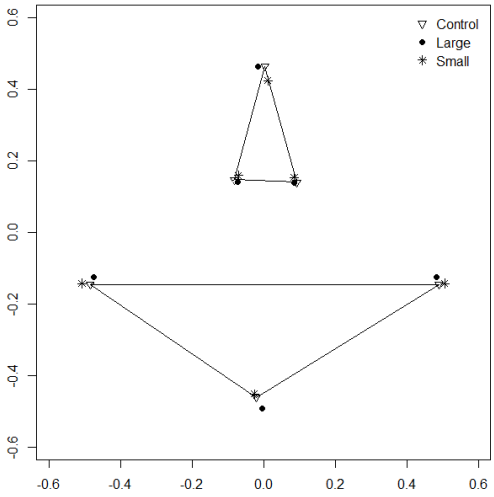


Figure 13 – Fit of mean shape of the landmark configurations using the generalized Procrustes analysis to $k+1=6$ landmarks. Illustration of the second thoracic vertebra T2 of mice.

In our study, we were interested at detecting the difference in the pre-shapes of triangles between the two groups in the two following contexts: (i) `control` vs. `large` and (ii) `control` vs. `small`. We have performed the study considering the cases 1 and 2 under high concentration. For the selection of the sample sizes 10 and 15 of the two groups, we made the resampling procedure with 1000 Monte Carlo experiments. We considered the resamples without replace of two groups.

Table 10 displays the rejection rates under null hypothesis of the proposed and F2 tests between the groups `control` and `large` at the case 1 and configuration A. Note that, for sample size 10 and 15, we verify the proposed distance-based tests detect the difference between mean shapes, while the F2 test does not reject the null hypothesis. But when considering the total sample size all tests present good results. Thus, we may partially conclude our proposals performed better than F2 in small samples, condition commonly found in SSA experiments.

In Table 28, we carried out the experiment for the comparison `control` vs. `large` at the case 2. Results indicate that the proposed tests have rejected the equality of mean

Table 10 – Rejection rates under H_0 for `control` vs. `large` at $\kappa_1 = \kappa_2$ and configuration A

Landmarks	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{F2}}$
(3,4,5)	10	10	0.047	0.048	0.048	0.047	0.048	0.032	0.164
	15	15	0.044	0.044	0.044	0.044	0.044	0.03	0.216
	30	23	0.04496	0.04492	0.04492	0.04496	0.04491	0.04917	0.0458

shapes as expected, with exception of the KL test at $n_1 = 30$ and $n_2 = 23$.

Table 11 – Rejection rates under H_0 for `control` vs. `large` at $\kappa_1 \neq \kappa_2$ and the configuration A

Landmarks	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{KL}}$
(3,4,5)	10	10	0.025	0.024	0.024	0.025	0.024	0	0
	15	15	0.035	0.034	0.034	0.035	0.033	0	0
	30	23	0.0311	0.03417	0.03417	0.0311	0.03519	0.04112	0.06243

Table 12 displays the rejection rates under null hypothesis for `control` vs. `small`, considering the configuration B and case 1. For two smallest sample sizes, the F2 test did not detect difference between the mean shapes, whilst the proposed tests identified the discrepancy. The increasing of the sample size tends to correct the F2 performance and all tests concludes what it is expected, there is difference between groups `control` and `small`.

Table 12 – Rejection rates under H_0 for group `control` vs. `small` at $\kappa_1 = \kappa_2$ and the configuration B

Landmarks	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{F2}}$
(1,2,6)	10	10	0.013	0.013	0.013	0.013	0.013	0.008	0.209
	15	15	0.025	0.025	0.025	0.025	0.025	0.021	0.322
	30	23	0.04997	0.04996	0.04996	0.04997	0	0	0.03165

Table 13 furnishes results for all proposed tests in `control` vs. `small` and the case 2. We recommend the H test because it has presented the empirical test sizes more closed to 5%.

In summary, considering the two configurations for the case 1, we noted the proposed tests outcomed the F2 test, what indicates our methodology as good tools to detect differences in complex spherical pre-shapes of triangles in the two-sample context. Taking

the case 2, the Hellinger statistic obtained better results, particularly for sample sizes $n_1 = 30$ and $n_2 = 23$.

Table 13 – Rejection rates under H_0 for group `control` vs. `small` at $\kappa_1 \neq \kappa_2$ and the configuration B

Landmarks	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$	$\hat{\alpha}_{S_{KL}}$
(1,2,6)	10	10	0.003	0.003	0.003	0.003	0.003	0	0
	15	15	0.006	0.004	0.004	0.006	0.004	0	0
	30	23	0.03763	0.04231	0.04231	0.03763	0.04388	0.05037	0.02209

3.5 CONCLUSION

This chapter has proposed four distance-based two-sample hypothesis tests for quantifying differences between triangle mean shapes. We have assumed the well-defined complex Watson ($\mathbb{C}W$) distribution for describing pre-shape data. We have furnished expressions for Rényi, Kullback-Leibler, Bhattacharyya and Hellinger distances on the $\mathbb{C}W$ model as well as have motivated their use in SSA for triangle complex spherical data. As a consequence from the use of proposed stochastic distances, we have provided pivotal quantities, which have been successfully employed to verify if two samples of triangle pre-shapes come from the same spherical probability measure on both low and high concentration degrees. The performance of our proposals has been quantified by means of Monte Carlo experiments and compared with the S_{F2} statistics test, well-defined tool for the $\mathbb{C}W$ model. Results have presented evidence that the S_R, S_B, S_H and S_{KL} tests may present estimated type I errors which are controlled. Moreover, estimated (in terms of rotation matrices) power functions of the proposed tests (with exception of the Kullback-Leibler version which is invariant under rotation) have obtained the good behavior, even on low concentration. Finally, we have illustrated the potentiality of these measures on real data. Our application has evaluated the difference between mean shapes of the two groups on second thoracic vertebrae T2 of mice. Results have favored three of our proposed tests, mainly in small sample size experiments.

4 DIVERGENCE-BASED PIVOTAL STATISTICS FOR PLANAR-SHAPE

4.1 Introduction

In this chapter, we propose new divergence-based two-sample tests in planar-shape. First we derive the Rényi, KL, Bhattacharyya and Hellinger distances for the $\mathbb{C}B$ distribution and, as a consequence, provide four relating hypothesis tests. Further we prove that the KL-based statistics is invariant under rotation and, therefore, it is able only to study homogeneity. Adopting as figures of merit both (i) test size and power and (ii) robustness, we make a Monte Carlo study to quantify the performance of our proposals, in comparison with other four SSA tests: Hotelling T^2, Goodall, James and lambda. Results indicate divergence-based tests have empirical test size as controlled (closed to adopted nominal level) as the literature ones; but, our proposals outperform the remainder in terms of both empirical test power and robustness. Finally, an application to real data is made in the O'Higgins and Dryden (1993) context. Results provide evidence in favor of all our proposals, with exception of the KL test (which has statistics invariant under rotation).

The remainder of this chapter is organized as follows. Section 5.1 approaches a survey of some of main planar-shape literature tests. In section 4.3 addresses the proposal of new divergence-based hypothesis tests. Section 4.4, we furnish numerical results from synthetic and actual experiments. Concluding remarks are presented in section 4.5. Subsequently, the Appendix B provides details about the derived results.

4.2 A survey about some hypothesis tests for planar-shape

Now some hypothesis tests (for mean shapes) which are well-defined in the SSA literature are discussed. Over the years, proposing statistical inference methods in planar-shape has been addressing by many papers. Lele and Richtsmeier (1991) have provided a test for shape difference based on a matrix analysis which was equipped by an Euclidean distance. Lele and III (1996) have addressed to test the equality of two shapes from

landmark configuration data when covariance matrices are unequal. Rohlf (2000b) has presented a survey about important SSA two-sample tests for shape equality as well as a comparative study of their empirical test powers. Micheas and Dey (2005b) have derived hypothesis tests and credible regions for average shape difference from a Bayesian viewpoint, assuming the complex elliptical shape distribution. The hypothesis tests which were proposed by Hotelling T^2, Goodall F, James (SEBER, 1984) and lambda (AMARAL; DRYDEN; WOOD, 2007) have aimed to check if there is shape difference between two independent random samples which are extracted from two-dimensional objects. In what follows, we discuss these four two-sample tests. The common hypotheses of these tests are

$$H_0 : [\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2] \,(= [\boldsymbol{\mu}_0]) \quad \text{vs.} \quad H_1 : [\boldsymbol{\mu}_1], [\boldsymbol{\mu}_2] \quad \text{unrestricted}, \tag{4.1}$$

where $[\boldsymbol{\mu}_0]$ is the common mean shape. The notation $[\boldsymbol{\mu}_j] = \{\exp(\mathrm{i}\upsilon)\boldsymbol{\mu}_j : 0 \leq \upsilon \leq 2\pi\}$ for $j = 1, 2$ represents the shape, corresponding to the modal pre-shape $\boldsymbol{\mu}_j$. The full Procrustes estimate of mean shape $[\widehat{\boldsymbol{\mu}}]$ can be defined as the eigenvector which corresponds to the largest eigenvalue of product matrix $\boldsymbol{S}$ (DRYDEN; MARDIA, 2016). Below, we describe, briefly, the exiting test present in literature.

a) Lambda statistics (AMARAL; DRYDEN; WOOD, 2007): Consider $\{\boldsymbol{X}_{ij}; j = 1, \ldots, n_i\}$ for $i = 1, \ldots, p$ be p independent samples at $\mathbb{C}^d$ and let $\widehat{\boldsymbol{\mu}}_i$ be the estimator for $\boldsymbol{\mu}_0$ based on the ith sample. Under the assumptions in Amaral, Dryden and Wood (2007) and choosing the matrix $\widehat{\boldsymbol{M}_i}$ for $i = 1, \ldots, p$, such as in this reference, we have that

$$n^{1/2}\, \widehat{\boldsymbol{M}_i}\, \boldsymbol{\mu}_0 \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{C}N_{d-1}(0, \boldsymbol{G}_i),$$

where

$\boldsymbol{G}_i$ has full rank and is a positive definite hermitian matrix. So, assuming

$$T_0(\boldsymbol{\mu}) \equiv 2\, \boldsymbol{\mu}^*\, \widehat{\boldsymbol{A}}_0\, \boldsymbol{\mu} \quad \text{and} \quad \widehat{\boldsymbol{A}}_0 = n \sum_{i=1}^{k} \widehat{\boldsymbol{M}}_i^*\, \widehat{\boldsymbol{G}}_i^{-1}\, \widehat{\boldsymbol{M}}_i \tag{4.2}$$

for $\widehat{\boldsymbol{G}}_i$ is the be a consistent estimator for $\boldsymbol{G}_i$, the lambda statistic – denoted by $\lambda_{\min}$ – is the smallest eigenvalue of $\widehat{\boldsymbol{A}}_0$ in (4.2) and $\widehat{\boldsymbol{\mu}}_0$ is the corresponding unit eigenvector. So,

$$\lambda_{\min} = \min_{\{\boldsymbol{\mu}:||\boldsymbol{\mu}||=1\}} [T_0(\boldsymbol{\mu})] = T_0(\widehat{\boldsymbol{\mu}}_0).$$

Under some conditions proposed in Fisher et al. (1996), one has that

$$\lambda_{\min} \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{2(p-1)(d-1)},$$

where p is the sample size d is the dimension of the space considered.

b) Hotelling T^2 statistic (AMARAL; DRYDEN; WOOD, 2007; DRYDEN; MARDIA, 2016): Let $\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_{ni}}$ for $i = 1, 2$ be two independent random samples which are drawn from population variables with mean shapes $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. We consider a two-sample test for data in the Procrustes tangent space. In this context, the pole corresponds to the overall pooled full Procrustes with mean shape $\widehat{\boldsymbol{\mu}}_i$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_1}$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{n_2}$ be the partial Procrustes tangent coordinates (with pole $\widehat{\boldsymbol{\mu}}_i$, $i = 1, 2$). A multivariate normal model is proposed in the tangent space, where $\boldsymbol{v}_i \sim \mathrm{N}(\boldsymbol{\xi}_1, \boldsymbol{\Sigma}_1)$ for $i = 1, 2, \ldots, n_1$, $\boldsymbol{w}_j \sim \mathrm{N}(\boldsymbol{\xi}_2, \boldsymbol{\Sigma}_2)$ for $j = 1, 2, \ldots, n_2$ and $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ are all mutually independent. Let $(\overline{\boldsymbol{v}}, \overline{\boldsymbol{w}})$ and $(\boldsymbol{S}_1, \boldsymbol{S}_2)$ denote the pairs of sample means and covariance matrices in each group. If the covariance matrices are assumed to be equal, then the squared Mahalanobis distance between $\overline{\boldsymbol{v}}$ and $\overline{\boldsymbol{w}}$ is $D^2 = (\bar{\boldsymbol{v}} - \bar{\boldsymbol{w}})^\top \boldsymbol{S}_{\boldsymbol{u}}^{-}(\bar{\boldsymbol{v}} - \bar{\boldsymbol{w}})$, where $\boldsymbol{S}_{\boldsymbol{u}} = (n_1 + n_2 - 2)^{-1}(n_1\boldsymbol{S}_{\boldsymbol{1}} + n_2\boldsymbol{S}_{\boldsymbol{2}})$ and $\boldsymbol{S}_{\boldsymbol{u}}^{-}$ is the Moore-Penrose generalized inverse of $\boldsymbol{S}_{\boldsymbol{u}}$. Under H_0, we have $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2$ and the two-sample Hotelling T^2 statistics is

$$F_H \;=\; \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)\,(n_1 + n_2 - 2)\,M} D^2,$$

where $M = 2p - 2$ is the dimension of the planar-shape space. This statistic follows the $F_{M,(n_1+n_2-M-1)}$ distribution under H_0 and the associated test is appropriate when variability is small and hence the tangent space approximation is reasonable. As decision rule, one rejects H_0 for large values of F_H.

c) James statistics (JAMES, 1954; SEBER, 1984): When covariance matrices in the previous item are not assumed to be equal, an alternative is to use the test statistic proposed by James (1954), given by

$$F_J \;=\; (\bar{\boldsymbol{v}} - \bar{\boldsymbol{w}})^\top \left(\frac{1}{n_1} S_{\boldsymbol{v}} + \frac{1}{n_2} S_{\boldsymbol{w}}\right)^{-} (\bar{\boldsymbol{v}} - \bar{\boldsymbol{w}}).$$

This statistics follows asymptotically the χ^2_M distribution, under H_0, regardless of whether or not $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are equal. In this case, we reject H_0 for large values of F_J.

d) Goodall statistics (GOODALL, 1991): If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ and we have the isotropic covariance structure $\boldsymbol{\Sigma} = 2\sigma^2\boldsymbol{I}$, then the statistic reduces to Goodall (1991) two-sample F statistic. Under H_0, the following statistic has the $F_{M,(n_1+n_2-2)M}$ distribution:

$$F_G = \left(\frac{n_1+n_2-2}{{n_1}^{-1}+{n_2}^{-1}}\right)\cdot\frac{d_F^2(\hat{\boldsymbol{\mu}}_1,\hat{\boldsymbol{\mu}}_2)}{\sum_{i=1}^{n_1} d_F^2(\boldsymbol{X}_{1i},\hat{\boldsymbol{\mu}}_1)+\sum_{j=1}^{n_2} d_F^2(\boldsymbol{X}_{2j},\hat{\boldsymbol{\mu}}_2)},$$

this result is valid for small σ and we reject H_0 for large values of F_G.

Now we discuss shortly assumptions of previous four tests. According to Amaral, Dryden and Wood (2007), the lambda statistic is designed for highly concentrated data.

The Goodall statistic (DRYDEN; MARDIA, 2016, p. 199) assumes the complex normal joint distribution on the landmark space, rotational symmetry and equal dispersion structure across populations. The Hotelling T^2 statistic (DRYDEN; MARDIA, 2016, p. 187) allows general dispersion structures and normality for the observations on the tangent space. The James statistic (SEBER, 1984, p.115) is a modified version of the Hotelling T^2 statistic, admitting different dispersion structures. The four tests we discussed previously will be compared.

4.3 New Divergence-Based Tests in Planar-Shape

Now we present closed-form expressions for the KL, Rényi, Bhattacharyya and Hellinger distances between $\mathbb{C}B$ probability measures.

Theorem 4.1. *Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_2)$ with pdfs given, respectively, by $f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)$. Then the KL distance is expressed by*

$$\begin{aligned} d_{KL}(\boldsymbol{A}_1,\boldsymbol{A}_2) = {} & \frac{1}{2}\Big[\sum_{j=1}^{k}\lambda_j^{(1)}\frac{1}{c(\boldsymbol{\Lambda}_{(1)})}\frac{\partial c(\boldsymbol{\Lambda}_{(1)})}{\partial\lambda_j^{(1)}} - \sum_{j=1}^{k}\lambda_j^{(2)}\frac{1}{c(\boldsymbol{\Lambda}_{(1)})}\frac{\partial c(\boldsymbol{\Lambda}_{(1)})}{\partial\lambda_j^{(1)}}\Big] \\ & + \frac{1}{2}\Big[\sum_{j=1}^{k}\lambda_j^{(2)}\frac{1}{c(\boldsymbol{\Lambda}_{(2)})}\frac{\partial c(\boldsymbol{\Lambda}_{(2)})}{\partial\lambda_j^{(2)}} - \sum_{j=1}^{k}\lambda_j^{(1)}\frac{1}{c(\boldsymbol{\Lambda}_{(2)})}\frac{\partial c(\boldsymbol{\Lambda}_{(2)})}{\partial\lambda_j^{(2)}}\Big], \end{aligned}$$

where $\boldsymbol{\Lambda}_{(q)} = \mathrm{diag}\{\lambda_1^{(q)},\lambda_2^{(q)},\ldots,\lambda_k^{(q)}\}$ is the diagonal matrix of eigenvalues of $\boldsymbol{A}_q$ for $q = 1, 2$. The expression for the first derivative of the $\mathbb{C}B$ normalization constant is

$$\frac{\partial c(\boldsymbol{\Lambda})}{\partial\lambda_j} = 2\pi^k\left[\sum_{l\neq j}^{k}\frac{a_l\exp(\lambda_l)}{(\lambda_l-\lambda_j)} + a_j\exp(\lambda_j) - a_jb_j\exp(\lambda_j)\right],$$

with $a_j = \prod_{i \neq j} \frac{1}{(\lambda_j - \lambda_i)}$ *and* $b_j = \sum_{i \neq j} \frac{1}{(\lambda_j - \lambda_i)}$.

Theorem 4.2. *Under conditions in Theorem 5.1 the Rényi distance* (d_R^β) *with order parameter* $\beta \in (0,1)$ *is*

$$d_R^\beta(\boldsymbol{A}_1, \boldsymbol{A}_2) = \frac{1}{\beta - 1} \log \left\{ \frac{c^{-\beta}(\boldsymbol{\Lambda}_1) c^{(\beta-1)}(\boldsymbol{\Lambda}_2) c(\Lambda_{R12}) \; + \; c^{-\beta}(\boldsymbol{\Lambda}_2) c^{(\beta-1)}(\boldsymbol{\Lambda}_1) c(\Lambda_{R21})}{2} \right\},$$

where $\boldsymbol{\Lambda}_i$ *is the matrix of eigenvalues of* $\boldsymbol{A}_i$ *for* $i = 1, 2$, $\boldsymbol{\Lambda}_{R12}$ *is the matrix of eigenvalues of* $[\beta \boldsymbol{A}_1 + (1-\beta)\boldsymbol{A}_2]$ *and* $\boldsymbol{\Lambda}_{R21}$ *is the matrix of eigenvalues of* $[\beta \boldsymbol{A}_2 + (1-\beta)\boldsymbol{A}_1]$.

The proof for the two previous theorems is in Appendix B. The following corollary holds from Theorem 4.2.

Corollary 4.1. *Under conditions in Theorem 4.2, Hellinger and the Bhattacharyya distances are*

$$d_H(\boldsymbol{A}_1, \boldsymbol{A}_2) = 1 - [c(\boldsymbol{\Lambda}_1) c(\boldsymbol{\Lambda}_2)]^{-1/2} c[(\boldsymbol{\Delta}_3)],$$

and

$$d_B(\boldsymbol{A}_1, \boldsymbol{A}_2) = \frac{1}{2} \log[c(\boldsymbol{\Lambda}_1) c(\boldsymbol{\Lambda}_2)] - \log[c(\boldsymbol{\Delta}_3)],$$

where $\boldsymbol{\Lambda}_1$ *and* $\boldsymbol{\Lambda}_2$ *are the matrices of eigenvalues of* $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$, *respectively, and* $\boldsymbol{\Delta}_3$ *is the matrix of eigenvalues of* $\frac{1}{2}(\boldsymbol{A}_1 + \boldsymbol{A}_2)$.

A proof of these results

can be found in Appendix B. According to Kent, Constable and Er (2004), the matrix parameters $\boldsymbol{A}$ and $\boldsymbol{A} + \Upsilon \boldsymbol{I}_k$, $\Upsilon \in \mathbb{R}$, indicate the $\mathbb{C}B$ same distribution. Hence, without loss of generality, we may shift the eigenvalues of $\boldsymbol{A}$ so that they are non positive with the largest one equals to 0. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k = 0$ denote the eigenvalues of $-\boldsymbol{A}$. This change is used for all simulation procedures in this chapter. Figure 14 illustrates the behavior of all distances proposed in Theorems 5.1 and 4.2 and Corollary 4.1. For this, we generate a $\mathbb{C}B$ distributed sample with size $n = 30$, vector of eigenvalues $\lambda = (200, 100, 0)$ and number of landmarks $k + 1 = 4$. We obtain as MLE the matrix

$$\boldsymbol{A} = \begin{pmatrix} 10.299737 + 0\text{i} & 35.028267 - 5.218285\text{i} & -0.272037 + 1.593317\text{i} \\ 35.028267 + 5.218285\text{i} & 90.690588 - 0\text{i} & 3.499143 - 1.834944\text{i} \\ -0.272037 - 1.593317\text{i} & 3.499143 + 1.834944\text{i} & 199.009675 - 0\text{i} \end{pmatrix},$$

which is used to illustrate the derived distances. The distances between $\boldsymbol{A}$ and $(1+\tau)\boldsymbol{A}$ for $|\tau| < 0.01$ are then plotted in Figure 14. For the Rényi distance, we assume the order parameter $\beta \in \{0.1, 0.3, 0.7, 0.9\}$. It is noticeable the following inequality holds:

$$d_R^{0.1} \leq d_B \approx d_H \leq d_R^{0.3} \leq d_R^{0.7} \leq d_R^{0.9} \leq d_{KL}.$$

Further, d_R^{β} tends to d_{KL} when $\beta \uparrow 1$.

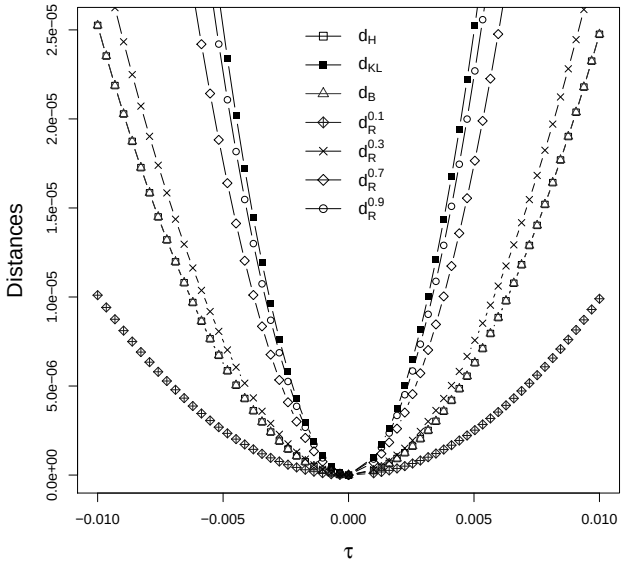


Figure 14 – Illustration of the Hellinger, Kullback-Leibler, Bhattacharyya and Rényi (with order $\beta \in \{0.1, 0.3, 0.7, 0.9\}$) distances. For graphical procedure we consider the configuration: sample size $n = 30$, vector eigenvalues $\lambda = (200, 100, 0)$ and number of landmarks $k + 1 = 4$. We compute the distances between $\boldsymbol{A}$ and $(1+\tau)\boldsymbol{A}$ with $|\tau| < 0.01$

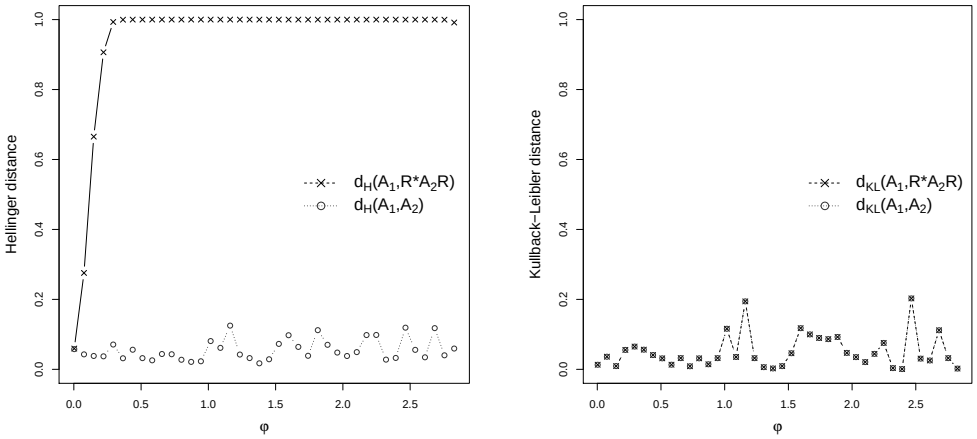
In contrast with the expressions for d_R^{β}, d_B and d_H, the KL distance depends only of eigenvalues of matrices in its arguments. This fact is linked with the property of rotation invariant, which is satisfied only by d_{KL}.

Proposition 4.1. *Let* $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_1)$ *and* $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_2)$. *Let* $\boldsymbol{R}$ *be a rotation matrix, then the Kullback-Leibler distance is invariant under scalar rotation, i.e.,*

$$d_{KL}(\boldsymbol{A}_1, \boldsymbol{A}_2) = d_{KL}(\boldsymbol{A}_1, \boldsymbol{R}^*\boldsymbol{A}_2\boldsymbol{R}).$$

Details about rotation matrix and the proof of Proposition 4.1 are presented in Appendix B. Figure 15 illustrates this property. In particular, we investigate the effect of rotation

(dependent of an angle, $\phi \in (0, 0.9\pi)$) over the KL and Hellinger distances. Large values angle imply high differences between $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, such as discussed by Amaral, Dryden and Wood (2007). Note that the Hellinger distance detects the increasing of variation; while the KL measure is not sensible to such variation.



(a) Hellinger distance (b) Kullback-Leibler distance

Figure 15 – Distances between two matrices with and without rotation in terms of the angle $\phi \in (0, 0.9\pi)$. For graphical procedure we take: eigenvalues vector $\lambda = (200, 100, 0)$, sample size $n = 30$ and number of landmarks $k + 1 = 4$

4.3.1 New hypothesis tests

Now, we are in position to present the proposed hypothesis tests. Consider $\widehat{\boldsymbol{A}_1}$ and $\widehat{\boldsymbol{A}_2}$ be MLEs for $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ based on two observed samples having sizes n_1 and n_2, respectively. We recall MLEs for eigenvectors of the matrix $\boldsymbol{A}_j$ are ones of $\boldsymbol{S}_j$ with $j = 1, 2$. Further, one should use an optimization non linear procedure for obtaining MLEs for eigenvalues of $\boldsymbol{A}_j$ and, as a consequence from the spectral decomposition, matrix MLEs of $\boldsymbol{A}_j$ are obtained.

Consider initially we are interested to test

$$H'_0 : \boldsymbol{A}_1 = \boldsymbol{A}_2 \quad vs. \quad H'_1 : \boldsymbol{A}_1 \neq \boldsymbol{A}_2. \tag{4.3}$$

Once the matrices $\boldsymbol{A}_j$ are hermitian, it follows from the Spectral Theorem (**??**) that $\boldsymbol{\Gamma}_j^* \boldsymbol{A}_j \boldsymbol{\Gamma}_j = \boldsymbol{\Lambda}_j$, where $\boldsymbol{\Gamma}_j$ is the matrix of eigenvectors of $\boldsymbol{A}_j$. Thus, the hypotheses in (4.3) can be rewritten as:

$$H_0 : \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 \quad vs. \quad H'_1 : \boldsymbol{A}_1 \neq \boldsymbol{A}_2.$$

From the SSA physical perspective, the rejection of H_0 may indicate three scenarios:

i) There is difference between diagonal matrices of eigenvalues under common eigenvectors taking as nuisance parameters; i.e,

$$H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 \quad vs. \quad H_1 : \mathbf{\Lambda}_1 \neq \mathbf{\Lambda}_2.$$

ii According to Mardia and Dryden (1999), if $\lambda_1^j = \cdots = \lambda_{k-1}^j$, for $j = 1, 2$, then the $\mathbb{C}B$ distribution collapses in the complex Watson $\mathbb{C}W$ model, which has as parameters the mean shape (say $\boldsymbol{\mu}$) and a concentration parameter (say κ). Assuming κ as a nuisance common parameter, (4.3) is replaced by

$$H_0'' : [\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2] \quad vs. \quad H_1'' : [\boldsymbol{\mu}_1], [\boldsymbol{\mu}_2] \text{ unrestricted.} \tag{4.4}$$

where $\boldsymbol{\mu}_i$ is the populacional mean shape associated to ith sample for $i = 1, 2$. The hypothesis 4.4 are often used in the SSA literature.

iii) There is difference between full matrices. In this case, the test is over 4.3 .

In this chapter, we focus on the conditions i) and ii) because they are related with tests which were discussed in Section 5.1. By the argument in condition ii) one has – similarly the $\mathbb{C}B$ model which extends the $\mathbb{C}W$ distribution – "$H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2$" generalizes "$H_0 : [\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2]$" under restrictions ii.1) $\lambda_1^j = \cdots = \lambda_{k-1}^j = \lambda_0^j$, ii.2) $\lambda_k^j = 0$ and iii.3) $\lambda_0^j = q(\boldsymbol{\mu}_j)$ and $q : \mathbb{C}\boldsymbol{S}^{k-1} \to \mathbb{R}_+$ is a mapping which is determined from the analytical relation between the $\mathbb{C}B$ and $\mathbb{C}W$ models.

Under the regularity conditions discussed in Salicrú et al. (1994), the following lemma holds.

Lemma 4.1. *Let $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\theta}_{11}, \ldots, \widehat{\theta}_{1M})$ and $\widehat{\boldsymbol{\theta}}_2 = (\widehat{\theta}_{21}, \ldots, \widehat{\theta}_{2M})$ be MLEs for parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ having $M_0 < M$ common components based on independent samples of sizes n_1 and n_2, respectively. If $\dfrac{n_1}{n_1 + n_2} \longrightarrow \delta \in (0, 1)$ when $n_1, n_2 \to \infty$ and $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, then*

$$S_\varphi^h(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \frac{2n_1n_2}{n_1 + n_2} \xrightarrow[n_1, n_2 \to \infty]{\mathcal{D}} \chi^2_{M - M_0},$$

where υ is the σ-finite measure on the measurable space $(\mathcal{X}, \mathcal{B})$, "$\xrightarrow[n_1, n_2 \to \infty]{\mathcal{D}}$" represents the convergence in distribution*, $D = \{KL, R, B, H\}$ and ν is a chosen value according to*

the used stochastic distance; i.e., $\nu = [h'(0)\varphi''(1)]^{-1} = \{1, 1/\beta, 4, 4\}$. *In what follows, we assume* $\mathcal{X} = \{1\}$ *and* $\upsilon(1) = 1$, *which yields*

$$\int_{\mathcal{X}} h'(0)\varphi''(1)\,\mathrm{d}\upsilon = h'(0)\varphi''(1)$$

(Salicrú et al., 1994, assumption adopted in Example 1).

Based on Lemma (4.1) (for the proof, see Salicrú et al. (1994, pp. 377-381)), the tests for the null hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ can be derived in the form of the following proposition.

Proposition 4.2. *Let* n_1 *and* n_2 *be large and* $S_D(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = s$, *then the null hypothesis* $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ *can be rejected at level* α *if the observed* s *is greater than the upper* α *quantile of* $\chi^2_{M-M_0}$ *distribution.*

It follows from Lemma 4.1 and Proposition 4.2 (taking $M_0 = 1$ due to identifiability condition of the $\mathbb{C}B$ model as discussed in Subsection 2.2) combined with Theorems 5.1 and 4.2 and Corollary 4.1, four new hypothesis tests for the $\mathbb{C}B$ distribution. The degrees of freedom of the pivotal statistics of proposed tests are associated to the condition i) that generalizes the case ii) (for which four literature tests are alternatives) and represent the quantity of the eigenvalues minus one (of restrictions $\lambda_k = 0$) to eliminate $\mathbb{C}B$ no identifiability.

a) Kullback-Leibler statistics - $S_{KL}(\cdot, \cdot)$:

$$S_{KL}(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) = \frac{2n_1n_2}{n_1 + n_2} d_{KL}(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) \underset{n_1,n_2\to\infty}{\overset{\mathcal{D}}{\to}} \chi^2_{k-1}.$$

b) Rényi statistics with order β - $S^{\beta}_R(\cdot, \cdot)$:

$$S^{\beta}_R(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) = \beta^{-1}\frac{2n_1n_2}{n_1 + n_2} d^{\beta}_R(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) \underset{n_1,n_2\to\infty}{\overset{\mathcal{D}}{\to}} \chi^2_{k-1}. \tag{4.5}$$

c) Bhattacharyya statistics - $S_B(\cdot, \cdot)$:

$$S_B(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) = \frac{8n_1n_2}{n_1 + n_2} d_B(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) \underset{n_1,n_2\to\infty}{\overset{\mathcal{D}}{\to}} \chi^2_{k-1}$$

d) Hellinger statistics - $S_H(\cdot, \cdot)$:

$$S_H(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) = \frac{8n_1n_2}{n_1 + n_2} d_H(\widehat{\boldsymbol{A}}_1, \widehat{\boldsymbol{A}}_2) \underset{n_1,n_2\to\infty}{\overset{\mathcal{D}}{\to}} \chi^2_{k-1}$$

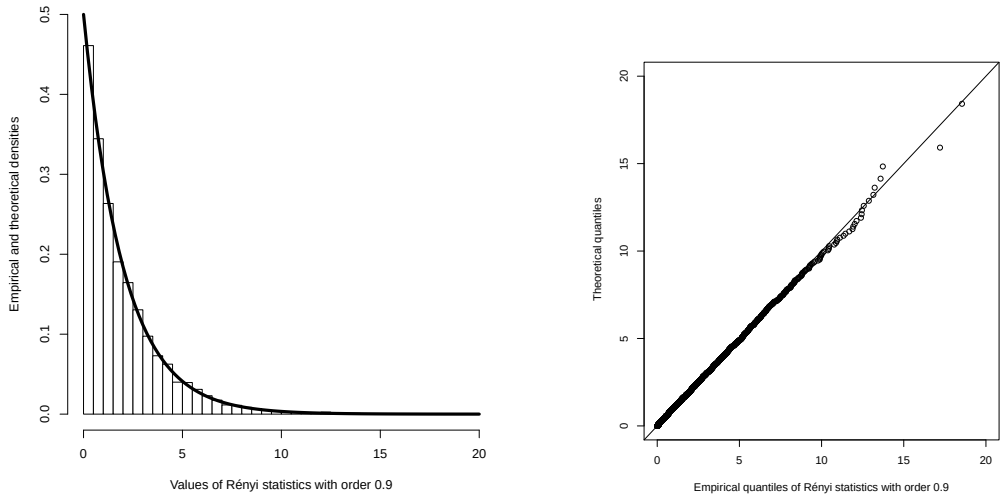
4.4 Numerical Results and Discussion

This section is organized in three parts. First the performance of well-defined and proposed tests is quantified in terms of estimated test sizes and powers for $\mathbb{C}B$ generated data. Second a robustness study for such tests is made. Finally an application to real data from a evolutionary biology experiment is designed and discussed. For numerical assessment, we use the literature tests which are implemented in the function resampletest on the package shapes (see <https://cran.r-project.org/web/packages/shapes/index.html>) of the software R.

4.4.1 Asymptotic behavior analysis of pivotal statistics from synthetic data

Here, we investigate how much considered tests approximate of what it is asymptotically expected. To that end we first present evidence proposed tests follow the chi-squared distribution with $k-1$ degrees of freedom. Once all pivotal statistics are (under null hypothesis) asymptotically equivalent, we display only results for the Rényi statistic at order $\beta = 0.9$, say $S_R^{0.9}$. Figures 16 exhibits results of a Monte Carlo simulation study, having 4000 replications to describe the limit distribution of $S_R^{0.9}$. In particular, we use outcomes from two $\mathbb{C}B$ distributed random samples having vector of eigenvalues $\lambda = \{200, 100, 0\}$, number of landmarks $k+1 = 4$ and sample sizes $n_1 = 70$ and $n_2 = 100$. Then, theoretical and empirical densities for values of $S_R^{0.9}$ are displayed in Figures 16a and 16b as well as the respective quantile-quantile plot (qq-plot). It is noticeable results favor derived statistics are in fact asymptotically chi-squared with $k-1$ degrees of freedom. In quantitative terms, the KS test has been applied to the array of observed statistics, yielding p-value 0.3513 as expected.

Now we present results of a test size study for proposed Kullback-Leibler, Rényi ($\beta \in \{0.1, 0.3, 0.7, 0.9\}$), Bhattacharyya and Hellinger tests. From the planar-shape literature as discussed by Dryden and Mardia (2016), it has noticed that the $\mathbb{C}B$ distribution imposes a hard estimation process which depends of the degree of concentration of data into the unit complex sphere. Then, we divide our study in two scenarios. We define the first and second scenarios as low and high concentrations, respectively. We use the term "high concentration" for data generated from the $\mathbb{C}B$ distribution at the eigenvalue vector



(a) Histogram of the Rényi statistic (b) Qqplot of the Rényi statistic

Figure 16 – Asymptotic behavior of the Rényi (with order $\beta = 0.9$) statistics. For graphical procedure we consider the configuration: eigenvalues vector $\lambda = \{200, 100, 0\}$, number of landmarks $k + 1 = 4$ and sample sizes $n_1 = 70$ and $n_2 = 100$

$\lambda = (\lambda_{11}, \lambda_{12}, 0)$, such that $\lambda_{11} > \lambda_{12}$ and $|\lambda_{11} - \lambda_{12}| \geq 100$; for otherwise, we denoted by "low concentration" Leu and Damien (2014) have used a similar setup. Further, in all the simulation procedure, we consider 4000 Monte Carlo replications with adopted significance level $\alpha = 0.05$.

Table 14 displays the empirical test sizes for the proposed tests under low concentration scenario. In general distance-based tests have demonstrated controlled empirical test sizes (i.e., good approximations to the adopted nominal level, 5%) for both small and large sample sizes. There is mild evidence at the case $\lambda = (8, 2, 0)$ that better rates are obtained as the sample size increases.

To compare rates in Table 14 with those of four well-defined planar-shape tests – say Hotelling T^2, James, Goodall and lambda – Table 15 shows rates of the latter. It has been observable the F_G test performed very poorly for all sample sizes and the $\lambda_{\min}$ test overestimated meaningfully the adopted nominal level for small sample sizes. About the $\lambda_{\min}$ statistic, results might be justified because this test has been designed for highly concentrated data. In general, the F_H and F_J tests have approached well 5% (i.e., have controlled estimated test sizes). These tests have similar rates and this can be justified because our studies assume samples which are generated under the equal covariance

Table 14 – Rejection rates under $H_0^{'}$ of distance-based (Rényi, Bhattacharyya, Hellinger and Kullback-Leibler) tests for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	$S_R^{0.1}$	$S_R^{0.3}$	$S_R^{0.7}$	$S_R^{0.9}$	S_B	S_H	S_{KL}
(10,5,0)	30	20	0.056	0.053	0.053	0.056	0.052	0.047	0.057
	30	50	0.052	0.05	0.05	0.052	0.052	0.048	0.052
	50	50	0.051	0.05	0.05	0.051	0.05	0.048	0.051
	70	100	0.048	0.048	0.048	0.048	0.048	0.047	0.049
	120	150	0.049	0.048	0.048	0.049	0.048	0.048	0.049
(8,2,0)	30	20	0.054	0.053	0.053	0.054	0.052	0.047	0.055
	30	50	0.058	0.058	0.058	0.058	0.058	0.056	0.058
	50	50	0.043	0.043	0.043	0.043	0.042	0.039	0.044
	70	100	0.05	0.05	0.05	0.05	0.05	0.049	0.051
	120	150	0.052	0.052	0.052	0.052	0.052	0.052	0.053

Table 15 – Rejection rates under $H_0^{''}$ of literature (Hotelling T^2, Goodall, James and lambda) tests for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	F_H	F_G	F_J	$\lambda_{\min}$
(10,5,0)	30	20	0.05	0.255	0.054	0.074
	30	50	0.055	0.271	0.058	0.08
	50	50	0.052	0.282	0.053	0.059
	70	100	0.057	0.284	0.058	0.066
	120	150	0.045	0.275	0.045	0.056
(8,2,0)	30	20	0.052	0.647	0.056	0.086
	30	50	0.053	0.663	0.054	0.086
	50	50	0.056	0.687	0.056	0.064
	70	100	0.052	0.685	0.051	0.062
	120	150	0.046	0.679	0.047	0.052

matrices.

Table 16 provides empirical test sizes for the proposed tests under high concentration. In general all tests have performed well, presenting rates which were closed to the adopted nominal level $\alpha = 0.05$. The worst results were obtained by S_B and S_H. In the Table 17, results have given evidence that F_H and F_J can furnish better rates than F_G and $\lambda_{\min}$, analogously to what happened on low concentration. In summary, for high concentration scenarios, results from Tables 16 and 17 have suggested that proposed tests can provide empirical test sizes as good as those of F_H and F_J.

Now we consider a study of test power (i.e., the rejection rates when samples come from

Table 16 – Rejection rates under $H_0^{'}$ of distance-based (Rényi, Bhattacharyya, Hellinger and Kullback-Leibler) tests for high concentration scenario. The adopted nominal level $\alpha = 5\%$

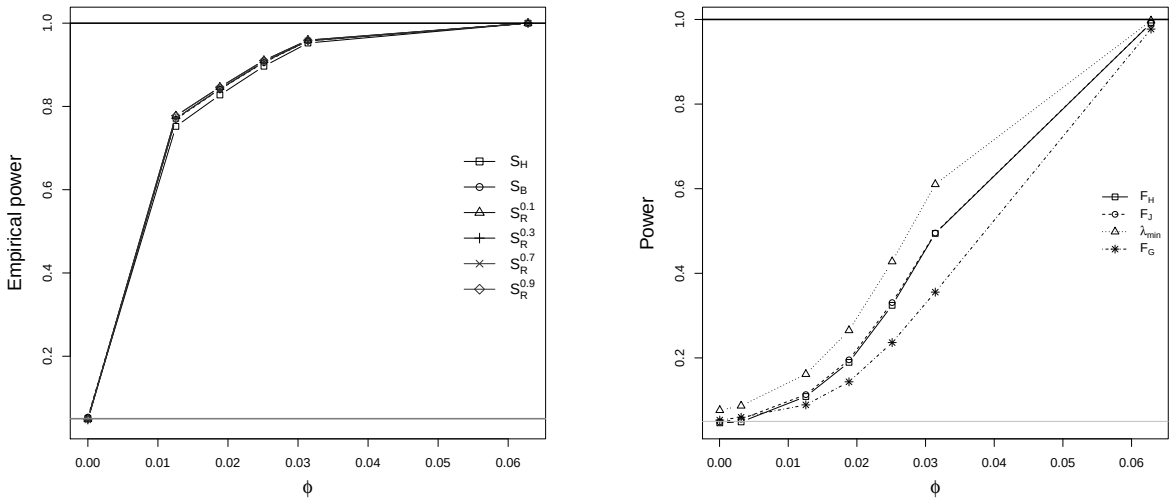
λ	n_1	n_2	$S_R^{0.1}$	$S_R^{0.3}$	$S_R^{0.7}$	$S_R^{0.9}$	S_B	S_H	S_{KL}
(200,100,0)	30	20	0.050	0.048	0.048	0.050	0.047	0.042	0.053
	30	50	0.056	0.054	0.054	0.056	0.054	0.051	0.058
	50	50	0.050	0.049	0.049	0.050	0.048	0.045	0.052
	70	100	0.051	0.051	0.051	0.051	0.050	0.048	0.052
	120	150	0.049	0.048	0.048	0.049	0.047	0.046	0.049
(300,150,0)	30	20	0.049	0.047	0.047	0.049	0.046	0.041	0.052
	30	50	0.055	0.054	0.054	0.055	0.054	0.051	0.058
	50	50	0.049	0.047	0.047	0.049	0.046	0.045	0.050
	70	100	0.051	0.050	0.050	0.051	0.049	0.048	0.052
	120	150	0.047	0.047	0.047	0.047	0.047	0.046	0.048

Table 17 – Rejection rates under $H_0^{''}$ of literature (Hotelling T^2, Goodall, James and lambda) tests for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	n_1	n_2	F_H	F_G	F_J	$\lambda_{\min}$
(200,100,0)	30	20	0.046	0.055	0.048	0.076
	30	50	0.046	0.061	0.05	0.078
	50	50	0.051	0.062	0.051	0.059
	70	100	0.049	0.058	0.048	0.058
	120	150	0.058	0.065	0.058	0.063
(300,150,0)	30	20	0.046	0.053	0.049	0.049
	30	50	0.046	0.059	0.05	0.078
	50	50	0.051	0.06	0.051	0.058
	70	100	0.049	0.058	0.048	0.058
	120	150	0.058	0.064	0.058	0.063

different distributions). For this, it is necessary to determine a mechanism of separation of H_0 and H_1. We employ a procedure called transformation of the null hypothesis, described by Amaral, Dryden and Wood (2007). In this procedure, we obtain a rotation matrix that – when applied to data – allows a sample to have the same mean shape. Figures 17a and 17b display several empirical power function cases for proposed and literature tests in terms of angle ϕ, respectively. Here the used sequence of angles for graphics of the power function has been defined in $\phi \in \{0, 0.001\pi, 0.004\pi, 0.006\pi, 0.008\pi, 0.01\pi, 0.02\pi\}$. Under high concentration scenarios, we considered the vector of eigenvalues $\lambda = (300, 150, 0)$ and sample sizes $n_1 = 30$ and $n_2 = 20$. We generated two samples under H_0 and

after transformed one of them to make H_0 false, pre-multiplying by the rotation matrix. Taking the estimated test power as comparison criterion, our proposals have performed meaningfully better than F_H, F_G, F_J and $\lambda_{\min}$ tests. Among the literature tests, the lambda test presented the highest power. For the proposed statistics, the Rényi (with $\beta \in \{0.1, 0.9\}$) tests provided the best rates among both proposed and literature tests.



(a) Power function for distance-based tests under high concentration scenario

(b) Power function for planar-shape literature tests under high concentration scenario

Figure 17 – Empirical power function for discussed tests under high concentration. For graphical configuration we take: eigenvalue of vector $\lambda = (300, 150, 0)$, sample sizes $n_1 = 30$ and $n_2 = 20$, number of landmarks $k + 1 = 4$ and the angle selected to compute the rotation matrix ϕ

Table 18 presents values for the power function of proposed distance-based tests under low concentration. Note that the power values tend to one as the sample size and angle increase as expected. Results of empirical power function under high concentration are given in the Table 19. Proposed tests have presented behavior better than those in lower concentration. It is expected because changes under high concentration are more apparent, implying in higher empirical power. Table 20 presents a power study for the literature tests under the low concentration scenarios. The lambda test has presented the best performance, but its growth was slower than divergence-based tests. Table 21 gives empirical power results under the high concentration scenario. Results pointed out all tests worked well. In summary, the proposed tests were meaningfully more powerful than literature tests, more expressive differences are given for low concentration and small

rotation. For high concentration, considering the angles $\phi_{power} = \{0.001\pi, 0.008\pi\}$, the proposed ones overcame the literature tests.

Table 18 – Rejection rates under $H_1^{'}$ of distance-based (Rényi, Bhattacharyya and Hellinger) tests for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	ϕ_{power}	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$
(10,5,0)	0.001π	30	50	0.67475	0.67300	0.67300	0.67475	0.67225	0.65975
		50	50	0.67850	0.6765	0.6765	0.67850	0.67650	0.66525
		70	100	0.68575	0.68525	0.68525	0.68575	0.6850	0.67700
	0.008π	30	50	0.70825	0.70550	0.70550	0.70825	0.70525	0.69450
		50	50	0.68775	0.6865	0.6865	0.68775	0.68625	0.67825
		70	100	0.69700	0.69625	0.69625	0.69700	0.6960	0.68950
	0.05π	30	50	0.92475	0.92425	0.92425	0.92475	0.92375	0.91850
		50	50	0.95975	0.9595	0.9595	0.95975	0.95925	0.95625
		70	100	0.98750	0.98750	0.98750	0.9875	0.9875	0.98725

Table 19 – Rejection rates under $H_1^{''}$ of distance-based (Rényi, Bhattacharyya and Hellinger) tests for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	ϕ_{power}	n_1	n_2	$\hat{\alpha}_{S_R^{0.1}}$	$\hat{\alpha}_{S_R^{0.3}}$	$\hat{\alpha}_{S_R^{0.7}}$	$\hat{\alpha}_{S_R^{0.9}}$	$\hat{\alpha}_{S_B}$	$\hat{\alpha}_{S_H}$
(200,100,0)	0.001π	30	50	0.6930	0.68875	0.68875	0.6930	0.68725	0.67200
		50	50	0.691	0.68675	0.68675	0.691	0.68525	0.6745
		70	100	0.67075	0.66775	0.66775	0.67075	0.66725	0.66125
	0.008π	30	50	0.8985	0.89725	0.89725	0.8985	0.89725	0.89275
		50	50	0.940	0.93975	0.93975	0.940	0.93975	0.9350
		70	100	0.98075	0.98075	0.98075	0.98075	0.98050	0.98000
	0.05π	30	50	1.00	1.00	1.00	1.00	1.00	1.00
		50	50	1.00	1.00	1.00	1.00	1.00	1.00
		70	100	1.00	1.00	1.00	1.00	1.00	1.00

Note that the KL test has not been considered in the previous discussion about power. According to Proposition 4.1, the KL distance is invariant under rotation and, therefore, the KL test is not adequate to test (4.1). However, this test may be useful to detect changes in eigenvalues of the $\mathbb{C}B$ parameters; meaning changes in the degree of concentration. To illustrate this, we also perform a second way of variation under the alternative hypothesis. We consider an additional simulation study to assess power in terms of change in eigenvalues of the $\mathbb{C}B$ parameter. To that end, we adopt the following setup: (i) two samples of $\mathbb{C}B$ generated data at the vectors of eigenvalues $\lambda = (300, 150, 0)$

Table 20 – Rejection rates under $H_1^{''}$ of distance-based (Hotelling T^2, James, Goodall and lambda) tests for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	ϕ_{power}	n_1	n_2	F_H	F_J	F_G	$\lambda_{\min}$
(10,5,0)	0.001π	30	50	0.05450	0.05425	0.27600	0.08125
		50	50	0.05150	0.05200	0.27975	0.06050
		70	100	0.05475	0.05500	0.27275	0.06475
	0.008π	30	50	0.05800	0.05600	0.27050	0.08650
		50	50	0.06500	0.06500	0.29400	0.07200
		70	100	0.06325	0.06200	0.30875	0.07675
	0.05π	30	50	0.41625	0.41250	0.66725	0.50925
		50	50	0.51850	0.52050	0.75425	0.57225
		70	100	0.79175	0.78950	0.91400	0.84300

Table 21 – Rejection rates under $H_1^{''}$ of distance-based (Hotelling T^2, James, Goodall and lambda) tests for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	ϕ_{power}	n_1	n_2	F_H	F_J	F_G	$\lambda_{\min}$
(200,100,0)	0.001π	30	50	0.05225	0.05250	0.06550	0.08300
		50	50	0.05750	0.05775	0.06375	0.0650
		70	100	0.05525	0.05350	0.06450	0.07100
	0.008π	30	50	0.35600	0.35475	0.25125	0.44625
		50	50	0.47500	0.47725	0.35625	0.51600
		70	100	0.70575	0.70800	0.54775	0.74975
	0.05π	30	50	1.00	1.00	1.00	1.00
		50	50	1.00	1.00	1.00	1.00
		70	100	1.00	1.00	1.00	1.00

and $(1+\varepsilon)\cdot\lambda$ such that $\varepsilon \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9, 1.0\}$, and (ii) number of landmarks $k+1 = 4$, the sample sizes $n_1 = 70$ and $n_2 = 100$ and 4000 Monte Carlo replicas. Figure 18 exhibits results of empirical powers. As literature tests are appropriate only for changes in mean shapes, we do not consider them in this study. We compare the performance of KL and Rényi (with order $\beta \in \{0.005, 0.9\}$) tests. Note that both KL and Rényi rejection rates have increased for one, but the Rényi test showed to be more powerful than the KL test. Furthermore, independently if we choose the order (β) close or away from 1, the behavior of the Rényi statistic is the same.

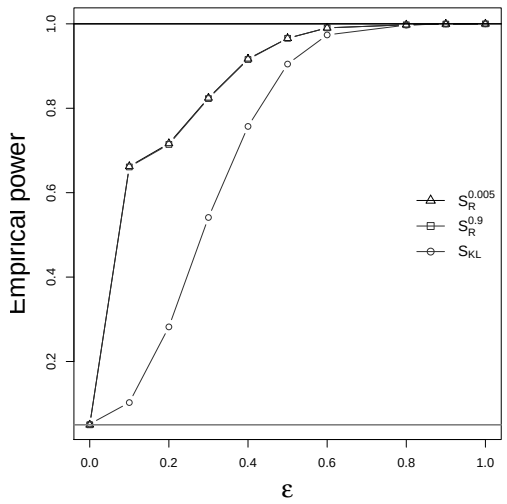


Figure 18 – Empirical power function for the proposed Rényi (with order $\beta \in \{0.005, 0.9\}$) and KL tests without rotation matrix. For simulation procedure we take: sample sizes $n_1 = 70$ and $n_2 = 100$, eigenvalues vector $\lambda = (300, 150, 0)$. To ensure that the procedure is under the alternative hypothesis we pre-multiply the eigenvalue vector of the second sample by $(1 + \varepsilon)$ such that $\varepsilon \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9, 1.0\}$

4.4.2 Robustness analysis from synthetic data

Beside comparing considered tests by their asymptotic performances, we also study how much they are robust before contaminated spherical samples. Specifically 4000 Monte Carlo replicas have been considered and we implemented the following procedures:

a) We generate two samples coming of the $\mathbb{C}B$ under high concentration scenario with eigenvalues vector $\lambda_1 = \{300, 150, 0\}$) with sample sizes $n_1 - 2$ and $n_2 - 2$;

b) We simulate one pair of the $\mathbb{C}B$ generated outcomes, weakly concentrated with eigenvalues vector $\lambda_2 = \{10, 5, 0\}$;

c) We concatenate the highly concentrated samples with the two observations obtained in the low concentration scenario.

Tables 22 and 23 approach empirical rejection rates under contaminated samples for (i) distance-based and (ii) planar-shape literature tests, respectively. Proposed tests have presented estimated test sizes which were near to 5%; while the performance of

remainder tests has been strongly affected by inserting points of low concentrated in highly concentrated samples.

Table 22 – Rejection rates under $H_0^{'}$ of distance-based (Rényi, Bhattacharyya, Hellinger and Kullback-Leibler) tests for contaminated samples. The adopted nominal level $\alpha = 5\%$

n_1	n_2	$S_R^{0.1}$	$S_R^{0.3}$	$S_R^{0.7}$	$S_R^{0.9}$	S_B	S_H	S_{KL}
30	50	0.051	0.049	0.049	0.051	0.048	0.044	0.053
70	100	0.043	0.042	0.042	0.043	0.041	0.04	0.043
120	80	0.051	0.05	0.05	0.051	0.05	0.049	0.052
140	200	0.049	0.048	0.048	0.049	0.048	0.048	0.049

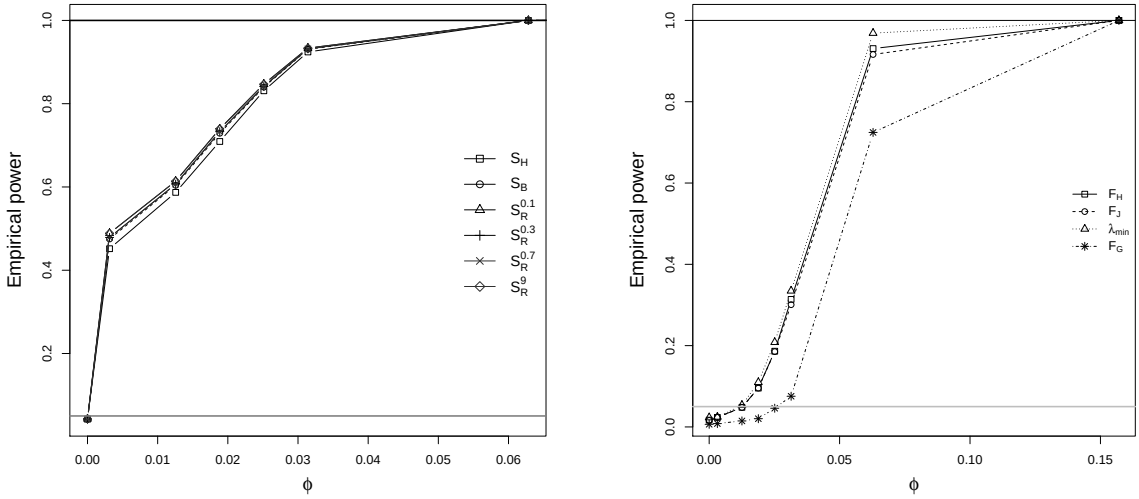
Table 23 – Rejection rates under $H_0^{''}$ of literature (Hotelling T^2, James, Goodall and lambda) tests for contaminated samples. The adopted nominal level $\alpha = 5\%$

n_1	n_2	F_H	F_G	F_J	$\lambda_{\min}$
30	50	0.015	0.003	0.015	0.015
70	100	0.017	0.011	0.016	0.022
120	80	0.024	0.02	0.024	0.027
140	200	0.035	0.036	0.034	0.044

To complement the previous discussion Figures 19 exhibits curves of power function for proposed and literature tests, respectively. For this numeric procedure we consider the angle $\phi \in \{0, 0.001\pi, 0.004\pi, 0.006\pi, 0.008\pi, 0.01\pi, 0.02\pi, 0.05\pi\}$ for computation of the rotation matrix and the sample sizes $n_1 = 30$ and $n_2 = 50$. It can note for the literature tests in the Figure 19b that the Goodall test showed worst robustness, while the lambda test presented the best behavior. For proposed tests in Figure 19a, the Hellinger test has been the less robust, comparatively those defined from the Rényi (with order $\beta \in \{0.1, 0.9\}$) test. In short, based in the results synthetic and real data, we indicate the Rényi for any order parameter and Bhattacharyya tests for detect difference between mean shapes.

4.4.3 Evolutionary biology data analysis

This section addresses an application in evolutionary biology. Specifically, the sexual dimorphism Oxnard (1983) is investigated in the species *Pan troglodytes* (chimpanzee), which is a member of the genus *Pan* and of the biological family *Hominidae* (humans, chimpanzees, gorillas and orangutans).



(a) Power function for distance-based tests with the angle ϕ

(b) Planar-shape literature tests with the angle ϕ

Figure 19 – Empirical power function of considered tests for contaminated samples with $n_1 = 30$ and $n_2 = 50$

This dataset was discussed by O'Higgins and Dryden (1993).Utilizing some methods for planar-shape, authors have investigated similarities and differences in cranial sexual dimorphism of hominoids.

Here we are interested whether samples of male and female chimpanzees skulls have different mean shapes.

The dataset we use has been constructed based on $k = 8$ landmarks in the midline of the cranium of 28 males and 26 females.

Amaral, Dryden and Wood (2007), Dryden and Mardia (2016) and O'Higgins (1989) have presented discussions about the dataset we consider.

Two-sample tests for equal mean shape are considered for three subsets of landmarks:

- Configuration A: all eight landmarks, denominated by whole skull region;
- Configuration B: five landmarks $(3, 4, 5, 6$ and $7)$ consisting of the face region and
- Configuration C: five landmarks $(1, 2, 3, 7,$ and $8)$ consisting of the braincase region.

The anatomical landmarks located in O'Higgins and Dryden (1993) by an expert biologist on cranium region of apes are showed in Figure 20.

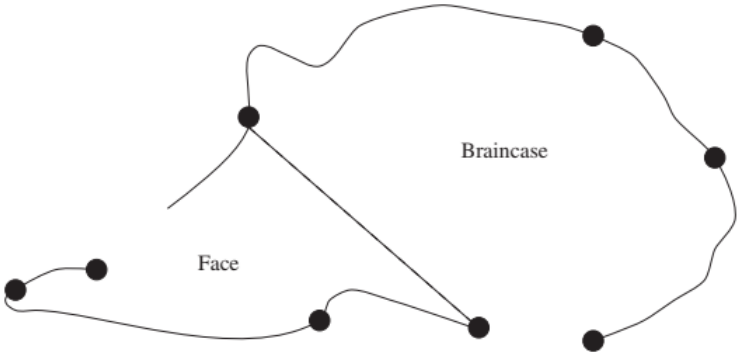
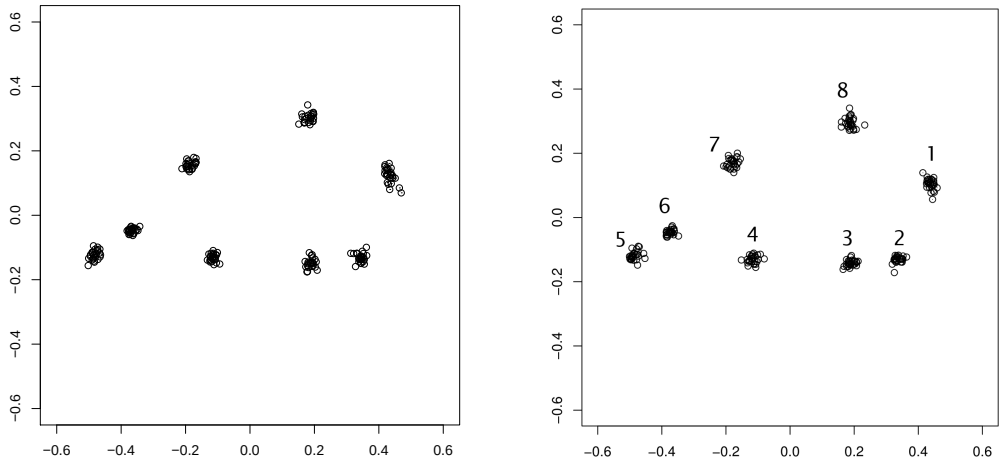


Figure 20 – Illustration of the configurations A, B and C. This image was taken from Dryden and Mardia (2016, pp. 19).



(a) Female of genus Pan. (b) Male of genus Pan.

Figure 21 – Samples of two-dimensional skulls having $k + 1 = 8$ landmarks, using full Procrustes fit.

In order to illustrate this dataset, Figure 21 presents the full Procrustes fit (DRYDEN; MARDIA, 2016) of the landmarks data for females and males chimpanzees. O'Higgins and Dryden (1993) have used several methods to study shape difference in the chimpanzees. According to their studies, considering over all landmarks, there was not statistically expressive shape difference between male and female chimpanzees. In contrast, an earlier study O'Higgins et al. (1990) found significant difference between male and female chimpanzees in terms of a number of facial measurements. Thus, as a take decision criterion, we have expected that – taking in account all landmarks – the tests do not detect changes in mean shapes, but in face braincase region the tests present significant difference. We

remember that the proposed pivotal statistics are developed from the $\mathbb{C}B$ distribution, while Hotelling's T^2, James and Goodall tests were obtained from the multivariate normal model and the lambda test has been deduced having as supposition the $\mathbb{C}N$ distribution.

Table 24 displays values of rejection rates for the planar-shape literature test in real data. For the Configuration A, results of James and Hotelling T^2 tests pointed out there is not evidence to reject the null hypothesis of equal mean shapes, as expected from O'Higgins and Dryden (1993). In contrast, Lambda and Goodall tests have rejected such null hypothesis. For Configuration B, James, Hotelling T^2 and Goodall tests have failed because they did not reject the null hypothesis in discordance with O'Higgins et al. (1990); but, the lambda test has provided evidence there is mean shape difference in terms of face between male and female chimpanzees skulls, as discussed by those authors. Finally, for Configuration C, only the Goodall test has confirmed there is mean shape difference in terms of braincase between male and female chimpanzees skulls.

Table 24 – Rejection rates under H_0'' of planar-shape literature tests in real data for configurations A,B and C. The adopted nominal level $\alpha = 5\%$

Region	Statistics	pvalue table
Whole skull		
$\lambda_{\min}$	30.577	0.0023
F_H	1.530	0.1527
F_J	23.4558	0.112
F_G	2.6021	0.0022
Face		
$\lambda_{\min}$	14.356	0.026
F_H	1.801	0.119
F_J	12.123	0.107
F_G	2.122	0.051
Braincase		
$\lambda_{\min}$	10.544	0.104
F_H	1.238	0.304
F_J	8.285	0.293
F_G	2.314	0.034

Table 25 gives values of rejection rates for proposed tests in real data. For configuration A, all tests have indicated there is not difference in mean shape in whole skull between male and female chimpanzees, what is in concordance with O'Higgins and Dryden (1993). For configurations B and C, with exception of the KL test, all tests have rejected the null

hypothesis of mean shape equality with strong significance as expected. We understand that the poor performance of the KL test can be explained by Proposition 4.1, for this application was focused on change of mean shape. In summary, this application seems to recommend Hellinger, Bhattacharyya and Rényi tests as good alternatives to detect difference in mean shape between samples in planar-shape.

Table 25 – Rejection rates under H_0' of distance-based tests in real data for configurations A,B and C. The adopted nominal level $\alpha = 5\%$

Region	Statistics	Distance	pvalue table
Whole skull			
$S_R^{0.1}$	8.874603	0.03291405	0.1807541
$S_R^{0.3}$	8.86156	0.09859703	0.1815148
$S_R^{0.7}$	8.86156	0.2300597	0.1815148
$S_R^{0.9}$	8.874603	0.2962265	0.1807541
S_H	8.503285	0.07884227	0.2034996
S_B	8.857225	0.082124	0.1817683
S_{KL}	0.02057014	0.0007629036	0.9999998
Face			
$S_R^{0.1}$	8.015961	0.02972953	0.04568298
$S_R^{0.3}$	8.014831	0.08917601	0.04570618
$S_R^{0.7}$	8.014831	0.2080774	0.04570618
$S_R^{0.9}$	8.015961	0.2675657	0.04568298
S_H	7.723926	0.07161607	0.05207556
S_B	8.014462	0.07430992	0.04571375
S_{KL}	0.0150608	0.0005585737	0.9995106
Braincase			
$S_R^{0.1}$	8.919959	0.03308226	0.03037423
$S_R^{0.3}$	8.905483	0.09908574	0.03057429
$S_R^{0.7}$	8.905483	0.2312001	0.03057429
$S_R^{0.9}$	8.919959	0.2977404	0.03037423
S_H	8.543345	0.07921371	0.03602082
S_B	8.900723	0.08252731	0.03064037
S_{KL}	0.2277743	0.008447675	0.9729857

4.5 Conclusions

In this chapter, we have proposed four divergence-based two-sample hypothesis tests for planar-shape. It has been assumed data follow the complex Bingham ($\mathbb{C}B$) distribution, which is a model widely used in Statistical Shape Analysis (SSA). First we have derived closed-form expressions for the Hellinger, Bhattacharyya, Kullback-Leibler (KL) and Rényi

distances between two $\mathbb{C}B$ probability measures. Practical insights about the use of them have been also provided. Second four divergence-based hypothesis tests for homogeneity in planar-shape have been proposed; particularly, Hellinger, Bhattacharyya and Rényi tests have also shown good ability to detect differences between mean shapes. We have proved the KL distance is invariant under rotation and, therefore, the KL test has worked well only for homogeneity. A Monte Carlo simulation study has been designed to quantify the performance (in terms of both test size and power and robustness) of new tests, in comparison with that due to other four SSA tests: Hotelling T^2, James, lambda and Goodall. Empirical test sizes and powers have provided evidence the proposed tests overcome the literature tests. An evaluation of robustness of the tests has also indicated the proposed tests have been more robust than literature ones. In applied terms, the Rényi, Hellinger and Bhattacharyya tests have obtained the best performance at assessing sexual dimorphism in chimpanzees skull. In general, numerical results from both simulated and real data have indicated divergence-based (in particular Bhattacharyya and Rényi) tests may outperform planar-shape literature tests, assuming the $\mathbb{C}B$ model as descriptor in planar-shape.

5 ENTROPY-BASED PIVOTAL STATISTICS FOR MULTI-SAMPLE PROBLEMS IN PLANAR-SHAPE

5.1 Introduction

In recent years, entropy classes have been used in statistical inference for various purposes; e.g., studying asymptotic properties of estimators alternative to the ML method Toma and Broniatowski (2011) and proposing robust hypothesis tests (GUL; ZOUBIR, 2016). Chen et al. (2016) have developed new insights about entropy behaviors as a multivariate variability measure. In this lastly perspective, the present chapter addresses the proposal of new entropy-based multi-sample tests for variability. To formulate these tests, we assume that pre-shape data are well-described by the $\mathbb{C}B$ distribution. We derive expressions for the Rényi and Shannon entropies for the $\mathbb{C}B$ and $\mathbb{C}W$ models. Moreover, these quantities are understood as entropy-based tests to assess if multiple spherical samples have the same degree of disorder, a kind of variability. To quantify the performance of proposed tests, we perform a Monte Carlo study, adopting empirical test sizes and powers as comparison criteria. An application to real data from the second thoracic vertebra T2 of mouses is done to assess the effect of body weight on the shape of mouse vertebra. Numerical results indicate that the test based on the optimized Rényi entropy may consist in a good tool to recognize similarity on planar-shape data.

The chapter unfolds as follows. Section 5.2 tackles the hypothesis tests for the equality of the entropies due to r independent populations based on information theory measures. Numerical results are displayed in Section 5.3. The conclusion remarks are presented in Section 5.4. Finally, proofs of the theoretical results are given in Appendix C.

5.2 New Entropy-Based Theoretical Results

In the following results, we derive new Shannon and Rényi entropies expressions under the supposition of the $\mathbb{C}B$ distribution for planar-shape .

Theorem 5.1. *Let* $\mathbf{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\mathbf{A})$. *For* $\beta \in \mathbb{R}_+ - \{1\}$, *we have that the Rényi and Shannon entropies are given by, respectively,*

$$H_R^{\beta}(\boldsymbol{A}) = \frac{1}{(1-\beta)}[\log c(\beta \boldsymbol{A}) - \beta \log c(\boldsymbol{\Lambda})] \tag{5.1}$$

and

$$H_S(\boldsymbol{A}) = \log c(\boldsymbol{\Lambda}) - \sum_{i=1}^{k-1} \lambda_i \frac{1}{c(\boldsymbol{\Lambda})} \frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_i}, \tag{5.2}$$

According to Pardo (2005), the Rényi entropy converges for the Shannon entropy when $\beta \to 1$. The corollary following enunciate this limit result for the $\mathbb{C}B$ distribution.

Corollary 5.1. *Let* $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$, *where* $\boldsymbol{A}$ *is a hermitian matrix. Then for all* $\beta \in \mathbb{R}_+ - \{1\}$, *we have*

$$\lim_{\beta \to 1} H_R^{\beta}(\mathbf{A}) = H_S(\mathbf{A}).$$

As discussed in Subsection 2.2.2, the $\mathbb{C}W$ is a special case of the $\mathbb{C}B$ model having two parameters: concentration κ and modal vector $\boldsymbol{\mu}$. The former parameter is the most explored in practice and larger values of $|\kappa|$ indicate more concentrated scenes. Now, first we derive expressions for Rényi and Shannon entropies in Theorem 5.2. Subsequently, we present the behavior of these entropies when κ increase.

Theorem 5.2. *Let* $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}, \kappa)$ *where* $\boldsymbol{\mu}$ *is modal vector of the spherical pre-shape and* κ *is concentration parameter. Then*

(a) the Rényi entropy is given by

$$H_R^{\beta}(\boldsymbol{\mu}, \kappa) = \frac{1}{(1-\beta)}\Big\{ -\beta \log[c_1(\kappa)] + \log[c_1(\beta\kappa)] \Big\}, \tag{5.3}$$

(b) and the Shannon entropy is expressed as

$$H_S(\boldsymbol{\mu}, \kappa) = \log[c_1(\kappa)] - \frac{\kappa}{k} \cdot \frac{{}_1F_1(2; k+1; \kappa)}{{}_1F_1(1; k; \kappa)}, \tag{5.4}$$

where $c_1(\cdot)$ *is the integrating constant.*

According to Mardia and Dryden (1999) and Dryden and Mardia (2016) when the concentration parameter $\kappa = 0$, the $\mathbb{C}W$ model collapses in the spherical uniform distribution whose density is the constant $2\pi^k/(k-1)!$. While when κ converge for infinite the $\mathbb{C}W$ model collapses in the $\mathbb{C}N$ distribution. Motivated by this asymptotic behavior, is relevant to evaluate the comportment of the entropy measures developed in the Theorem 5.2 when the concentration parameter κ converge for zero and infinite. In the next result we prove that, when the concentration parameter of the $\mathbb{C}W$ model is close to zero, the Rényi and Shannon entropies converge to the pdf of the spherical uniform distribution. We also evaluate the asymptotic behavior of the entropy measures when κ converges to infinity, but the resulting limit expression does not exist.

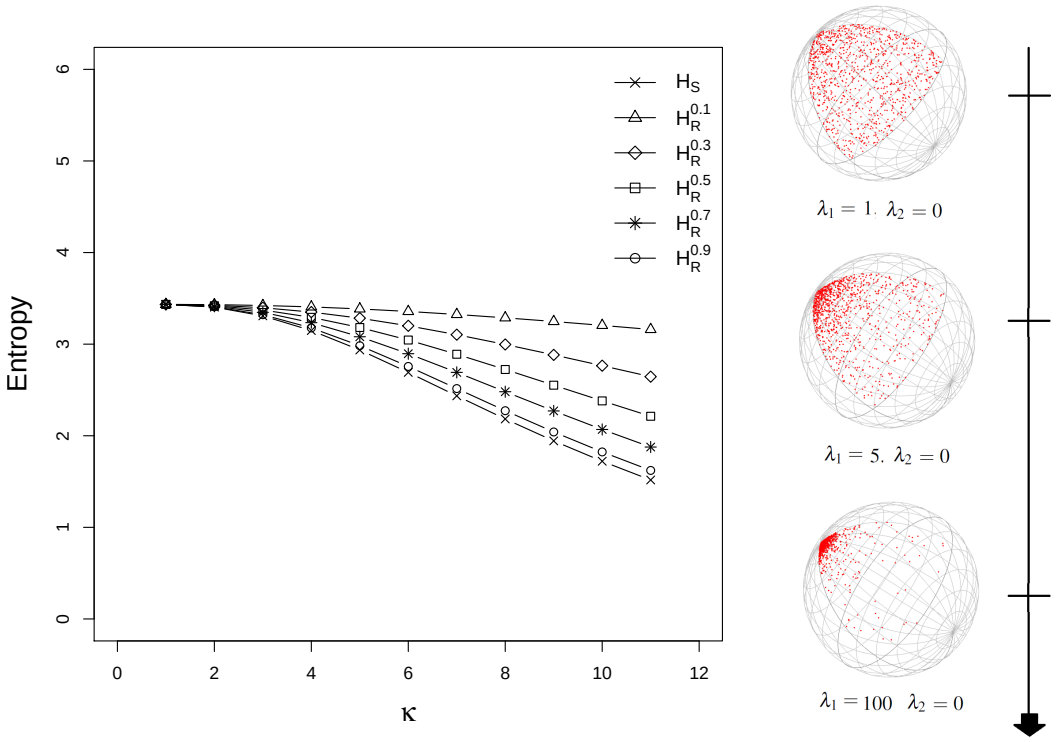
Corollary 5.2. *Let* $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\mu}, \kappa)$. *Whereas* $\boldsymbol{\mu}$ *fixed and* $\beta \in \mathbb{R}_+ - \{1\}$ *arbitrary. If* $\kappa \to 0$ *in the expressions* (5.3) *and* (5.4) *respectively, then*

$$\lim_{\kappa \to 0} H_R^{\beta}(\boldsymbol{\mu}, \kappa) = \lim_{\kappa \to 0} H_S(\boldsymbol{\mu}, \kappa) = \left[\frac{2\pi^k}{(k-1)!}\right].$$

The proof of all results above are given in the Appendix C.

Now we briefly investigate the behavior of the Shannon and Rényi entropies when increasing the concentration parameter. Consider the Rényi entropy measure order $\beta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, κ varying from 0 to 10 and landmarks number $k+1 = 5$ (case which will be analyzed in our simulation study). Figure 22a displays the Shannon and Rényi entropies curves. Note that when κ is very close to zero, the values of entropies are similar, as expected from the Corollary 5.2. Moreover, the behavior of measures indicates more concentrated data tend to have smaller values of entropy. To illustrate the last statement, Figure 22b shows the plot of three samples drawn from the $\mathbb{C}W$ distribution. As discussed by Mardia (1989) and Dryden and Mardia (2016), these data are pre-shapes of triangles, which can be represented either $\mathbb{C}\boldsymbol{S}^1$ or $\boldsymbol{S}^3$, being the real pre-shape sphere. In this case, the $\mathbb{C}W$ and $\mathbb{C}B$ models are similar. Specifically, we assume they have parameters $(\lambda_1, \lambda_2) = (\kappa, 0)$; i.e., the value of λ_1 given $\lambda_2 = 0$ and $\boldsymbol{\mu} = [0, 1]^\top$ is equal to κ by the identity $\boldsymbol{A} = \kappa\,(\boldsymbol{I}_2 - \boldsymbol{\mu}\boldsymbol{\mu}^*)$. It is noticeable the increasing of κ causes the concentration high degree in pre-shape data and, as a consequence, one has small

values of entropy may represent most concentrated spherical phenomena. This conclusion is important because it simplifies the behavior of complex vector data on the unit sphere to a real scale quantity, whose distribution and associated quantiles are more easy studied.



(a) Entropy curves (b) Pre-shape samples

Figure 22 – Illustration for both (i) the values of Shannon and Rényi entropies when $\beta \uparrow 1$ and κ increases and (ii) behavior of triangles pre-shapes over the unit sphere.

5.2.1 Pivotal statistics

In this subsection, we present some preliminary results that support the hypothesis test based on the Rényi and Shannon entropies. These tests are substantiated in Salicrú et al. (1994) and have been highly considered by Arellano-Valle, Contreras-Reyes and Stehlík (2017). They have established KL divergences between each generalized skew-normal distribution and the normal one in terms of their negentropies in order to develop hypothesis testing for normality.

The next Lemma was proposed and proved by Pardo (2005) and is assumed that the statistical space $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}$ satisfies some standard regularity conditions.

Lemma 5.1. *Let $\widehat{\boldsymbol{\theta}}$ be the ML estimator of $\boldsymbol{\theta}$ based on a random sample of size n drawn from a model with pdf $f(\boldsymbol{x};\boldsymbol{\theta})$. Assume the regularity conditions (mentioned before) are valid, $\varphi_{\mathcal{M}} \in C^1([0,\infty))$ (i.e. $\varphi_{\mathcal{M}}$ is derivable continuously in the interval $[0,\infty)$) and there exist a measurable and υ-integrable function $F(x)$ such that*

$$\left|\varphi'_{\mathcal{M}}(f(\boldsymbol{x};\boldsymbol{\theta}))\frac{\partial f(\boldsymbol{x};\boldsymbol{\theta})}{\partial\theta_j}\right| < F(\boldsymbol{x}), \quad j=1,\ldots,k;$$

where $\mathcal{M}=\{R,S\}$ denotes the choose between Rényi and Shannon measures, respectively, and

$$\varphi_{\mathcal{M}}(f(\boldsymbol{x};\boldsymbol{\theta})) = \begin{cases} -f(\boldsymbol{x};\boldsymbol{\theta})\log[f(\boldsymbol{x};\boldsymbol{\theta})] & , \quad \mathcal{M}=S \\ [f(\boldsymbol{x};\boldsymbol{\theta})]^{\beta} & , \ \mathcal{M}=R \ \text{ and } \ \beta\in\mathbb{R}_+-\{1\}. \end{cases}$$

Then

$$\sqrt{n}[H_{\mathcal{M}}(\widehat{\boldsymbol{\theta}}) - H_{\mathcal{M}}(\boldsymbol{\theta})] \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathrm{N}(0,\sigma^2_{\mathcal{M}}(\boldsymbol{\theta})),$$

where $\mathrm{N}(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, provided

$$\sigma^2_{\mathcal{M}}(\boldsymbol{\theta}) = \boldsymbol{T}_{\mathcal{M}}^{\top}\boldsymbol{I}_F(\boldsymbol{\theta})^{-1}\boldsymbol{T}_{\mathcal{M}} \tag{5.5}$$

and $\boldsymbol{T}_{\mathcal{M}} = [t_{\mathcal{M}_1}, t_{\mathcal{M}_2},\ldots,t_{\mathcal{M}_k}]^{\top}$ such that

$$t_{\mathcal{M}_j} = \frac{\partial H_{\mathcal{M}}(\theta)}{\partial\theta_j}, \quad j=\{1,2,\ldots,k\}. \tag{5.6}$$

Based on the Lemma 5.1, in order to find the expression for the variance with respect to each considered entropy in this thesis, we first determine the components of $\boldsymbol{T}_{\mathcal{M}}$ in the following proposition.

Proposition 5.1. *Let $\boldsymbol{z}\sim\mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$ with the Shannon entropy given in (5.2), we have that the vector $\boldsymbol{T}_S=[t_{S_1},t_{S_2},\ldots,t_{S_p}]^{\top}$ in (5.5) is given by*

$$t_{S_i} = \lambda_i\left[\frac{1}{c(\boldsymbol{\Lambda})^2}\left(\frac{\partial c(\boldsymbol{\Lambda})}{\partial\lambda_i}\right)^2 - \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial^2 c(\boldsymbol{\Lambda})}{\partial\lambda_i^2}\right], \quad \textit{for} \quad i=1,\ldots,k,$$

where $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\ldots,\lambda_{k-1})$ is the eigenvalues matrix of $\boldsymbol{A}$. With respect to the Rényi entropy in (5.1), we have the vector $\boldsymbol{T}_R=[t_{R_1},t_{R_2},\ldots,t_{R_p}]^{\top}$ such that

$$t_{R_i} = \frac{1}{(1-\beta)}\left[\frac{1}{c(\beta\boldsymbol{\Lambda})}\frac{\partial c(\beta\boldsymbol{\Lambda})}{\partial\lambda_j} - \frac{\beta}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial\boldsymbol{\Lambda}_i}\right],$$

where $\beta\in\mathbb{R}_+-\{1\}$.

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ such that $\boldsymbol{x}_j \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_j)$, $\boldsymbol{A}_j$ be parameters hermitian matrices for $j = 1, 2, \ldots, r$ and r is the number of under-study populations. Consider that we are interested at testing the null hypothesis

$$H_0 : H_{\mathcal{M}}(\boldsymbol{A}_1) = H_{\mathcal{M}}(\boldsymbol{A}_2) = \ldots = H_{\mathcal{M}}(\boldsymbol{A}_r) = D_0$$

versus $H_1 : \exists j, k \in \{1, \ldots, r\}$ such that $H_{\mathcal{M}}(\boldsymbol{A}_j) \neq H_{\mathcal{M}}(\boldsymbol{A}_k)$. Let $\widehat{\boldsymbol{A}_j}$ be the ML estimator for $\boldsymbol{A}_j$ based on a random sample of size n_j drawn from $\boldsymbol{x}_j$, for $j = 1, \ldots, r$. It follows from Lemma 5.1 that

$$\frac{\sqrt{n_j}(H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j}) - D_0)}{\sigma_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathrm{N}(0, 1)$$

for $j = 1, 2, \ldots, r$ and $\mathcal{M} \in \{R, S\}$. So

$$\sum_{j=1}^{r} \frac{n_j(H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j}) - D_0)^2}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} \underset{n_j\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \chi^2_r.$$

As D_0 is an unknown constant in practice, Pardo (2005) have rewritten this statistics as

$$\sum_{j=1}^{r} \frac{n_j(H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j}) - D_0)^2}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} = \sum_{j=1}^{r} \frac{n_j(H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j}) - \overline{D})^2}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} + \sum_{j=1}^{r} \frac{n_j(\overline{D} - D_0)^2}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} \tag{5.7}$$

where

$$\overline{D} = \left[\sum_{j=1}^{r} \frac{n_j}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})}\right]^{-1} \sum_{j=1}^{r} \frac{n_j H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})}.$$

Pardo (2005) has shown that the second parcel of the right side of (5.7) is distributed as a chi-squared with degree of freedom one. Thus,

$$S_{\mathcal{M}}(\widehat{\boldsymbol{A}_1}, \widehat{\boldsymbol{A}_2} \ldots, \widehat{\boldsymbol{A}_r}) = \sum_{j=1}^{r} \frac{n_j(H_{\mathcal{M}}(\widehat{\boldsymbol{A}_j}) - \overline{D})^2}{\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})} \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \chi^2_{r-1}. \tag{5.8}$$

The amount $\sigma^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_j})$ is obtained from the Lemma 5.1 and Proposition 5.1. Thus, the following decision rule ensues: the null hypothesis is rejected if

$$S_{\mathcal{M}}(\widehat{\boldsymbol{A}_1}, \mathbf{A}_2 \ldots, \widehat{\boldsymbol{A}_r}) > \chi^2_{(r-1),\alpha},$$

where $\chi^2_{(r-1),\alpha}$ denotes the $100\%(1-\alpha)$ percentile of the chi-squared distribution with $r-1$ (r is the number of population) degrees of freedom, fixing the significance level at α.

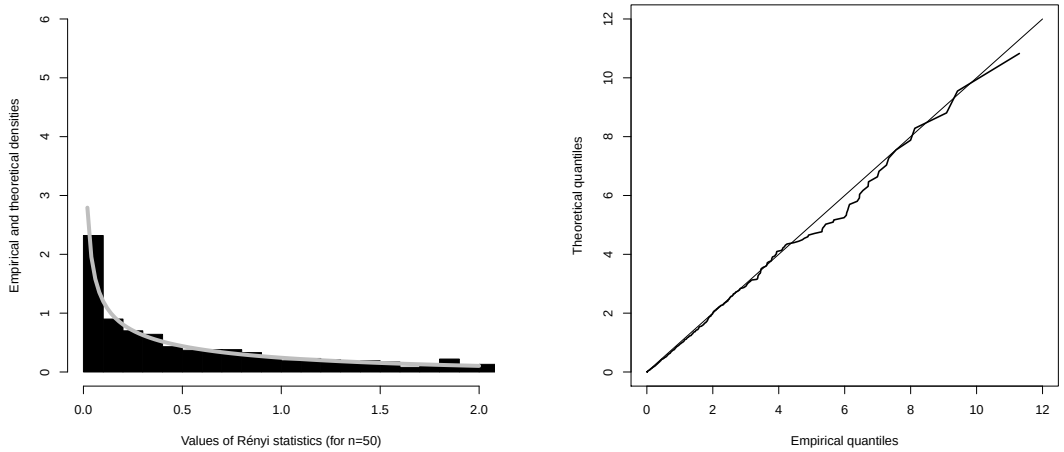
5.3 Numerical Results

Numerical results related to the proposed tests are presented in this section. First, their performance is evaluated for generated $\mathbb{C}B$ data. To that end, estimated sizes and powers are considered to compare the proposed methods. After our proposals are applied to actual data, the dataset mice available in the software R. All results were computed using functions of the package `shapes` (for more details, see <https://cran.r-project.org/web/packages/shapes/index.html>).

5.3.1 Simulation study

Initially, the asymptotic behavior of the proposed pivotal statistics is investigated by means of a short pilot study. It was performed with 5000 replicas and, over each one, were computed Shannon and Rényi statistics (with order 0.48, justified subsequently) in (5.8) for two-samples ($r = 2$) with size $n = 50$ and highly concentrated spherical behavior. Figures 23 and 24 display theoretical and empirical densities for Rényi and Shannon statistics, respectively, as well as the respective quantile-quantile plots (qq-plots). It was observable that the Shannon statistic failed to follow the chi-squared distribution, while the Rényi version at the order 0.48 was successful. The KS test was employed to the array of observed $S_R^{0.48}$ statistics, yielding p-value 0.2669; i.e., the hypothesis that the observed array comes from the chi-square distribution is not rejected to $\alpha \leq 5\%$. In order to illustrate that the Shannon test achieves the chi-squared behavior in some sample size, Figure 24 also includes the sample sizes $n = 100$ and 200. However, the future discussion focuses on sizes smaller that 100, which is a real demand for the use of morphometry tools in biological experiments.

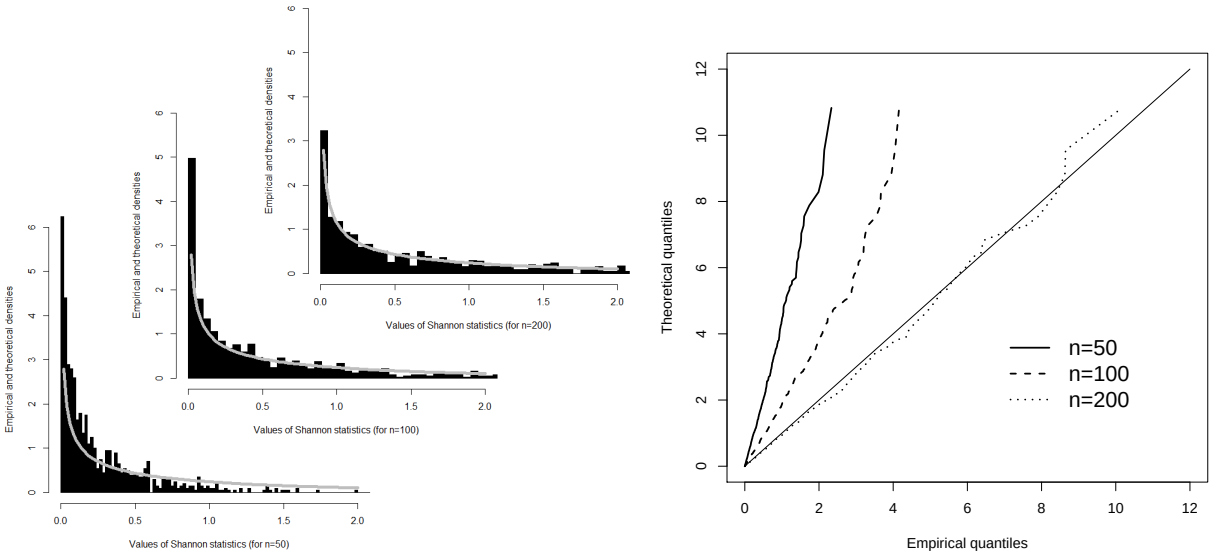
From this study, although the Shannon test did not work well, it is noticed that there is an order, say β_{otm}, that can optimize the Rényi test in the sense of proximity to the adopted nominal level. Now, we are in position of making a deeper study to confirm the existence of an order, β_{otm}, for S_R^{β}, that becomes this statistic adequate on small and moderate sample sizes. We suspect initially that the optimized value depends of the choice of the all quantities involved in tests proposed in Section 4.3: number of landmarks ($k + 1 = 5$), sample sizes $(30, 50, 70, 100, 150)$, number of populations ($r \in \{2, 3, 5, 6\}$) and



(a) Histogram of the Rényi statistic with order $\beta_{otm} = 0.48$

(b) Qq-plot of the Rényi statistics with order $\beta_{otm} = 0.48$

Figure 23 – Histogram and qq-plot of the Rényi statistics under high concentration for samples coming of the $\mathbb{C}B$ model. For simulation procedure we consider: eigenvalues vector $\lambda = (99, 42, 1, 0)$, sample sizes $n_1 = n_2 = n = 50$, populacional number r=2 and $k + 1 = 5$ landmarks



(a) Histogram of the Shannon statistic

(b) Qq-plot of the Shannon statistic

Figure 24 – Histogram and qq-plot of the Shannon statistics under high concentration for samples coming of the $\mathbb{C}B$ model. For simulation procedure we consider: eigenvalues vector $\lambda = (99, 42, 1, 0)$, sample sizes $n_1 = n_2 = n = 50, 100, 200$, populacional number r=2 and $k + 1 = 5$ landmarks.

two concentration degrees (low and high). The performance of S_S and S_R^β (with order parameter $\beta \in \{0.5, 0.9, \beta_{otm}\}$) tests was measured under 5000 Monte Carlo replicas.

Table 26 – Rejection rates under H_0 of entropy-based tests for low concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	r	n	S_S	$S_R^{0.5}$	$S_R^{0.9}$	$S_R^{\beta_{otm}}$	β_{otm}
(5,2,1,0)	2	30	0.2244	0.0022	0.0066	0.0492	1.9
		50	0.3704	0.0024	0.0084	0.046	1.57
		70	0.4372	0.0016	0.0072	0.0488	1.6
		100	0.5692	0.017	0.0352	0.048	1.3
	3	30	0.243	0.0028	0.0078	0.0474	1.8
		50	0.4138	0.0014	0.0034	0.0496	1.7
		70	0.5014	0.0022	0.0066	0.0464	1.65
		100	0.6184	0.0206	0.0364	0.0476	1.25
	5	30	0.9866	0.9156	0.9134	0.0506	0.028
		50	0.995	0.9814	0.9792	0.0476	0.041
		70	0.9944	0.9838	0.986	0.0452	0.048
		100	0.9858	0.9722	0.9764	0.051	0.05
	6	30	0.9886	0.9262	0.924	0.0426	0.025
		50	0.9964	0.9764	0.9748	0.051	0.042
		70	0.9962	0.9858	0.984	0.0506	0.048
		100	0.9858	0.9736	0.9726	0.0512	0.049

Tables 26 and 27 present values for the test size under the low and high concentration scenarios. For high concentration, it was observed some cases on which the non optimized S_R^β test have presented good results: $r = 2, 3$, $\beta \in \{0.5, 0.9\}$ and small size. However, in general, the test based in the Shannon entropy together with those in terms of the Rényi measure with order 0.5 and 0.9 have shown unsatisfactory results. But, for both concentration scenarios, we have found a way to optimize β so that the rejection rates under the null hypothesis have achieved a good result. We use the methodology described on Algorithm 3 to determine an acceptable β, β_{otm}.

Algorithm 3 Procedure to obtain β_{otm} with minimizes the rejection rates under the null hypothesis

1: Given r samples supposed to come from the same population having the support over the unit sphere, obtain B sets with r vectors of size n.
2: Define a grid of values for β in S_R^β as wide as possible and after determine an array with the rejection rates in terms of values of the adopted grid.
3: Choose β_{otm} as the value of β which minimizes the rejection rates in the array of the previous item.

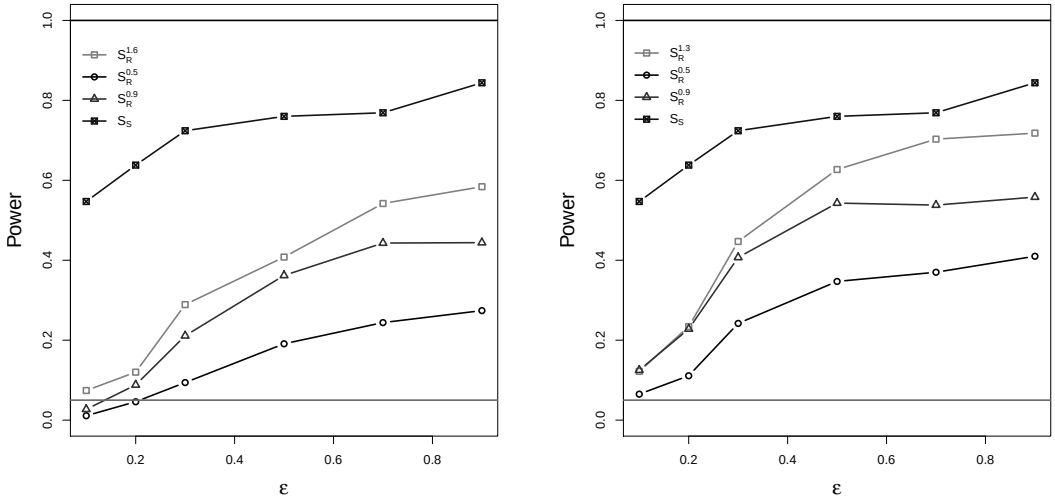
Table 27 – Rejection rates under H_1 of entropy-based tests for high concentration scenario. The adopted nominal level $\alpha = 5\%$

λ	r	n	S_S	$S_R^{0.5}$	$S_R^{0.9}$	$S_R^{\beta_{otm}}$	β_{otm}
(99,42,1,0)	2	30	0	0.015	0.0064	0.0478	0.18
		50	0	0.0482	0.0318	0.0498	0.48
		70	2e-04	0.0918	0.0704	0.0608	1.68
		100	0.001	0.1642	0.1308	0.125	1.22
	3	30	0	0.0036	0.0016	0.0448	0.12
		50	0	0.029	0.0186	0.0498	0.35
		70	0	0.0768	0.0538	0.0524	0.95
		100	2e-04	0.1454	0.1056	0.0468	1.36
	5	30	0	0.2556	0.0074	0.0462	4.4
		50	0	0.7354	0.081	0.0554	2.3
		70	4e-04	0.9448	0.2404	0.0496	1.8
		100	0.0084	0.9958	0.4992	0.0492	1.55
	6	30	0	0.2676	0.008	0.0456	0.71
		50	0	0.7676	0.0848	0.0462	0.99
		70	2e-04	0.9568	0.2506	0.0488	1.15
		100	0.008	0.9968	0.5244	0.053	1.6

In order to illustrate the discriminating capability of the Rényi test with the optimized order parameter comparatively to the remainder ones, we have quantified their power functions for some cases. Figure 25a and 25b shows the power function performance for low concentrations with sample sizes $n = 70$ and $n = 100$, respectively. It was considered the landmarks number $k+1 = 5$, 5000 Monte Carlo replicas, the sample size $n = \{70, 100\}$ and population number $r = 2$. In the simulation procedure, we take the eigenvalue vector $\boldsymbol{\lambda} = (5, 2, 1, 0)$ to define the null hypothesis to tests: S_R^{β} with $\beta \in \{0.5, 0.9, 1.3, 1.6\}$ ($\beta_{otm} = 1.3$ for $n = 100$ and $\beta_{otm} = 1.6$ for $n = 70$, according to Table 26) and S_S.

From this, we generate the first sample with $\boldsymbol{\lambda}$ and the second sample with $\underline{\boldsymbol{\lambda}} = (5 \cdot (1-\epsilon)^{2.5}, 2 \cdot (1-\epsilon)^{2.5}, 1, 0)$ where $\epsilon = [0.10, 0.20, 0.30, 0.50, 0.70, 0.90]$. Note that the Shannon test provided a power bigger than the Rényi test. However, as discussed, the S_S empirical test size showed unacceptable before considered sample sizes. On the other hand, among the possible Rényi tests, the $S_R^{\beta_{otm}}$ power outperformed ones due to the remainder tests. Further, the Rényi test power increases with the grow of the sample size.

Figure 26 displays the power function for high concentration scenario. For this, we consider $\boldsymbol{\lambda} = (99, 42, 1, 0)$ to specify the null hypothesis of tests: S_R^{β} with $\beta = (0.5, 0.9, 1.68)$ and S_S. The choose $\beta = 1.68$ is aligned with $n = 70$ in Table 27. Here, we consider only



(a) Power function under low concentration with sample size $n_1 = n_2 = 70$.

(b) Power function under low concentration with sample size $n_1 = n_2 = 100$.

Figure 25 – Empirical power function under low concentration. For this numeric procedure we consider: eigenvalue vector $\boldsymbol{\lambda} = (5, 2, 1, 0)$, order parameter $\beta \in \{0.5, 0.9, 1.3, 1.6\}$ for the Rényi statistic, populacional number $r = 2$ and $k + 1 = 5$ landmarks

one sample size for the power assessment on high concentration because tests performed accurately at $n = 70$. To that end, we generated the first sample with $\boldsymbol{\lambda}$ and the second with $\underline{\boldsymbol{\lambda}} = (99 \cdot (1 - \epsilon), 42 \cdot (1 - \epsilon), 1)$ such that $\epsilon = [0.1, 0.2, 0.3, 0.4]^\top$. We note the performance of the Rényi test power was better than for the Shannon test. In this case, the empirical power of the $S_R^{\beta_{otm}}$ test was the smallest in the Rényi class, but all these tests performed very well and almost similarly.

Observe that in this scenario for all orders β in S_R^β the power function approaches of 1, that is in the high concentration the Rényi statistic is very powerful. In general, best results were obtained for larger sample on low and high concentration scenario as expected.

5.3.2 Application to real data

We employ the derived tests to identify patterns over a database of measurements on the second thoracic vertebra T2 of mice, presented by Dryden and Mardia (2016) and previously discussed in Chapter 3. We wish to quantify how the body weight influences the shape of the mouse T2 vertebrae in the variability context. We work with three groups

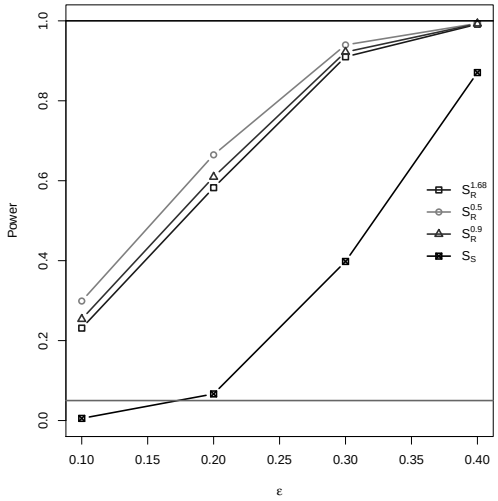


Figure 26 – Empirical power function under high concentration. For this numeric procedure we consider: eigenvalue vector $\boldsymbol{\lambda} = (99, 42, 1, 0)$, order parameter $\beta \in \{0.5, 0.9, 1.6\}$ for the Rényi statistic,sample size $n_1 = n_2 = n = 70$ populacional number $r = 2$ and $k + 1 = 5$ landmarks

of mice: `control` (with 30 elements), `large` (23) and `small` (23). Six landmarks were taken in each individual by a procedure described by Mardia and Dryden (1989a).

The `control` group contains unselected mice, while the `large` and `small` groups are formed by mice which were selected having large and small body weight, respectively. This database has been used by several works, it follows some instances of its use:

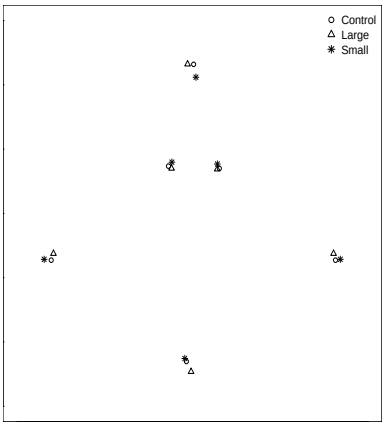
a. Kent (1994) has illustrated the use of principal components in shape analysis equipped with the Kendall's and Bookstein's coordinates.

b. Dryden and Mardia (1991) have discussed about biological aspects of three groups of mice. They have aimed both estimating and testing differences between shapes in two-dimensions over Bookstein's coordinates.

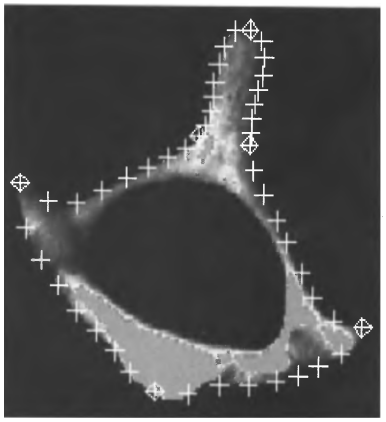
c. Mardia and Dryden (1989b) have investigated shape changes among groups and considered possible geometrical descriptions of first (T1) and second (T2) thoracic vertebrae. The bones they used were subset from the larger study detailed by Falconer (1973).

Figure 27a presents the full Procrustes fit of landmarks data concerning with the three groups of mice. For each group we obtain the mean shapes of the fitted data and plot

the $k + 1 = 6$ landmarks referring to the three groups (`control`, `large` and `small`). This graph was made using the function procrustes2d on the package shapes of the software R.



(a) A vertebra T2 from mouse which were registered using the mean shape of full Procrustes fit with six landmarks.



(b) Grey level image of a T2 mouse vertebra with six mathematical landmarks and 42 pseudo-landmarks (+). This figure was taken from Dryden and Mardia (1998), pp. 4

Figure 27 – Illustration of the second thoracic vertebra T2 of the group mice.

Table 10 displays average values of proposed test statistics ($\overline{S}_{\mathcal{M}}$), their p-values and estimated variances ($\widehat{\sigma}^2_{\mathcal{M}}(\widehat{\boldsymbol{A}_i})$) for various samples size. Before real data, we have verified the FIM in (5.6) was singular and, consequently, the use of (5.5) was not possible. To overcome this issue, we obtained $\sigma^2_{\mathcal{M}}$ in (5.5) by means of the non parametric bootstrap methodology. Initially, in order to construct a random framework from a real dataset we performed 750 resampling with replacement over each considered group. Since the implementation had some problems, an alternative method was used to compute the estimated asymptotic variance. After, we estimated the variance ($\sigma^2_{\mathcal{M}}$) using the following expression

$$\widehat{\sigma}^2_{\mathcal{M}} = \frac{1}{750}\sum_{j=1}^{750}(H^{(j)}_{\mathcal{M}} - \overline{H_{\mathcal{M}}})^2,$$

where $H^{(j)}_{\mathcal{M}}$ is the entropy $\mathcal{M} = \{S, R\}$ which was estimated over the jth bootstrap sample and $\overline{H_{\mathcal{M}}}$ is the bootstrap average of estimated entropies.

Table 28 shows values for the same quantities of Table 10 for high concentration. From both tables, the test based in the Rényi entropy rejected the null hypothesis adopting $\alpha \leq 5\%$, independent of the considered scenario. On the other hand, the test based on the Shannon entropy never rejects H_0. With respect the difference in pairs of entropies in the

set {`small`,`large`,`control`}, it is expected there are differences in pairs (`small`,`control`) and (`large`,`control`). Thus, these results seem to recommend in practice the Rényi test may be more adequate to detect the variability in planar pre-shapes over the complex unitary sphere.

Table 28 – Rejection rates under H_0, values of proposed statistics and their variances based on real data for high concentration. The adopted nominal level $\alpha = 5\%$

Tests	$\{n_1, n_2, n_3\}$	$\overline{S}_{\mathcal{M}}$	**pvalue table**	$\widehat{\sigma}^2_{\mathcal{M}}(\widehat{\boldsymbol{A}}_1)$	$\widehat{\sigma}^2_{\mathcal{M}}(\widehat{\boldsymbol{A}}_2)$	$\widehat{\sigma}^2_{\mathcal{M}}(\widehat{\boldsymbol{A}}_3)$
S_S	$\{10, 10, 10\}$	0.06285198	0.9690627	0.2868608	0.1290481	0.2586201
$S_R^{0.00775}$		6.233742	0.04429555	0.1017824	0.06037482	0.07393719
S_S	$\{15, 15, 15\}$	0.1191108	0.9421833	0.1383523	0.0916418	0.1015166
$S_R^{0.00746}$		6.057827	0.04836815	0.05435652	0.04583011	0.04742249
S_S	$\{30, 23, 23\}$	3.543456	0.1700389	0.02320577	0.06870133	0.1096711
$S_R^{0.0068}$		11.80529	0.00273221	0.01091856	0.02620905	0.04358041

5.4 Conclusions

This chapter has presented two entropy-based r-samples hypothesis tests for checking common variability in planar-shape data. The complex Bingham ($\mathbb{C}B$) model is assumed for the data. We have provided expressions for the $\mathbb{C}B$ Shannon and Rényi entropies. Some properties of these measures were also explored; such as new expressions for the complex Watson ($\mathbb{C}W$) particular model and limit relations between those entropies. Geometrical essays have indicated that the interpretation of entropy as a variability measure is aligned with the degree of concentration of planar pre-shapes over the unit sphere. Finally, entropy-based hypothesis tests for variability in planar-shape were proposed as well. The performance of our proposals has been quantified by means of a Monte Carlo simulation study. The Rényi entropy test with optimized order parameter obtained the best empirical test size in all the cases. Among possible tests to be defined from the Rényi class, that last achieved either lesser (but closed to) empirical power than remainder in the high concentration scenarios or expressively higher for low concentration. Finally, we illustrated the potentiality of these measures over real data to evaluate the effect of body weight over the shape of mouse vertebrae. The Rényi (in particular, with β optimized order) test got better performance than one in terms of the Shannon entropy.

6 CONCLUSION

In this chapter, we present some concluding remarks and future works that we intend to do in the course of this thesis.

6.1 Concluding Remarks

In this manuscript, as first research chapter, has presented four divergence-based two-sample hypothesis tests for checking the difference between mean triangle shape. The complex Watson ($\mathbb{C}W$) distribution was assumed for the data. We have furnished expressions for Rényi, Kullback-Leibler, Bhattacharyya and Hellinger distances measure on the $\mathbb{C}W$ model and illustrated its behavior. From these stochastic distances, we provide pivotal statistic and applied to triangle shapes data. The performance of our proposals has been quantified by means of a Monte Carlo simulation study and we compare our results with a literature statistics test proposed by Mardia and Dryden (1999). The largest differential of the our methodology is to be able applied the tests in the low concentration scenario for large samples and when the κ parameter in two-sample of the $\mathbb{C}W$ model were considered distinct. For the high concentration scenario, we have provided evidence in favor of the our proposals illustrated for both empirical type I and II errors study. Finally, was illustrated the performance of these measures over real data to evaluate the difference between mean shape of the two groups on second thoracic T2 of the mouse vertebrae.

In the second research topic, we have derived expressions for distance measures of Rényi, Hellinger, Bhattacharyya and Kullback-Leibler under $\mathbb{C}B$ distribution. After that, we have shown new pivotal statistics based in this information theory measures for two samples of planar-shape. Several simulation procedures were performed investigated the size and power of the proposed tests and compared them with the literature tests. Results indicated, under the assumption of $\mathbb{C}B$ model, that proposed tests presented the best performance. We made a brief numerical evaluation with respect the robustness of the proposed and literature tests. The results showed that the Rényi, Bhattacharyya, Hellinger and Kullback-Leibler tests can be considered more robust than literature tests.

Potentiality of our proposals have been illustrated experiments with synthetic and real data in planar-shape.

In the third topic, we have derived expressions for Shannon and Rényi entropies measures under the $\mathbb{C}B$ and $\mathbb{C}W$ models. From this, we have presented two tests for r-populations based on entropy measures in planar-shape. We have illustrated the discriminatory capability of these measures using on both synthetic and real data. Numerical results have pointed out that the tests based on the Rényi entropy with optimized β is indicated to detect changes in variability on the pre-shape vectors on complex sphere.

6.2 Future Research

This thesis has also raised several points, which require deeper studies. Some of them are presented below:

(i) to propose a nonparametric version of the divergence-based hypothesis tests of the Chapter 4 using the bootstrap methodology and permutation tests;

(ii) to develops new classifiers for shape data based on distances measures derived in the Chapter 3 and 4;

(iii) to propose a new regression model with response variable following the $\mathbb{C}B$ distribution. It, will be developed a method for detecting outliers in term over influential measures based on the pivotal statistics: Rényi, Bhattacharyya, Hellinger and Kullback-Leibler distances under the $\mathbb{C}B$ model;

(iv) to derive a method in optimization process to estimate the order parameter β in the hypothesis test based on the Rényi entropy;

(v) to provide new estimation procedure for the $\mathbb{C}B$ parameters based on the maximum entropy principle and/or the ordinary entropy method;

(vi) to furnish a procedure for generating random numbers from the complex-valued $\mathbb{C}W$ distribution own;

and

(vii) to develop new hypothesis tests based on Hotelling-Lawley trace statistic for the $\mathbb{C}B$ model.

REFERENCES

ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. [S.l.]: Courier Corporation, 1964.

AMARAL, G. A.; DRYDEN, I.; WOOD, A. T. Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. *Journal of the American Statistical Association*, Taylor & Francis, v. 102, p. 695–707, 2007.

AMARAL, G. J. et al. Bootstrap confidence regions for the planar mean shape. *Journal of Statistical Planning and Inference*, Elsevier, v. 140, p. 3026–3034, 2010.

AMARAL, G. J. a.; FLORÉZ, O. P. r.; CYSNEIROS, F. J. A. Graphical and numerical methods for detecting influential observations in complex Bingham data. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 42, p. 1801–1814, 2013.

ARELLANO-VALLE, R. B.; CONTRERAS-REYES, J. E.; STEHLÍK, M. Generalized Skew-Normal Negentropy and Its Application to Fish Condition Factor Time Series. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 19, p. 528, 2017.

BARATPOUR, S.; AHMADI, J.; ARGHAMI, N. R. Characterizations based on Rényi entropy of order statistics and record values. *Journal of Statistical Planning and Inference*, Elsevier, v. 138, p. 2544–2551, 2008.

BARTLE, R. G.; SHERBERT, D. R. *Introduction to Real Analysis*. [S.l.]: Wiley, 1982. (Matemáticas (Limusa)).

BASSEVILLE, M. Divergence measures for statistical data processing – an annotated bibliography. *Signal Processing*, Elsevier, v. 93, p. 621–633, 2013.

BHATTACHARYA, R.; PATRANGENARU, V. Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds: II. *Annals of Statistics*, JSTOR, p. 1225–1259, 2005.

BHATTACHARYA, R.; PATRANGENARU, V. et al. Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds: I. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 31, p. 1–29, 2003.

BILLINGSLEY, P. *Probability and Measure*. [S.l.]: John Wiley & Sons, 2008.

BINGHAM, C. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, JSTOR, v. 2, p. 1201–1225, 1974.

BLATT, D.; HERO, A. O. On Tests for Global Maximum of the Log-Likelihood Function. *IEEE Transactions on Information Theory*, IEEE Institute of Electrical and Electronics, v. 53, p. 2510, 2007.

BOOKSTEIN, F. L. A statistical method for biological shape comparisons. *Journal of Theoretical Biology*, Elsevier, v. 107, p. 475–520, 1984.

BOOKSTEIN, F. L. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, JSTOR, p. 181–222, 1986.

BOOKSTEIN, F. L. Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, Elsevier, v. 58, p. 313–365, 1996.

BROMBIN, C. et al. *Parametric and Nonparametric Inference for Statistical Dynamic Shape Analysis with Applications*. [S.l.]: Springer, 2016.

CHEN, B. et al. Insights into Entropy as a Measure of Multivariate Variability. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 18, p. 196, 2016.

CHUNG, J. et al. Measures of distance between probability distributions. *Journal of Mathematical Analysis and Applications*, Elsevier BV, v. 138, p. 280–292, 1989.

COLAK, C. et al. Detecting the Shape Differences of the Corpus Callosum in Behçet's Disease by Statistical Shape Analysis. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, Wiley Online Library, v. 294, p. 870–874, 2011.

CONTRERAS-REYES, J. E. Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy–tailed distributions. *Physica A: Statistical Mechanics and Its Applications*, Elsevier, v. 395, p. 200–208, 2014.

CONTRERAS-REYES, J. E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. *Physica A: Statistical Mechanics and Its Applications*, Elsevier, v. 433, p. 84–91, 2015.

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

CSISZÁR, I. Information Type Measures of Difference of Probability Distributions and Indirect Observations. *Studia Scientiarum Mathematicarum Hungarica*, v. 2, p. 299–318, 1967.

DARLING, D. A. The Kolmogorov-Smirnov, Cramer-Von Mises Tests. *The Annals of Mathematical Statistics*, JSTOR, v. 28, p. 823–838, 1957.

DEZA, M. M.; DEZA, E. *Encyclopedia of Distances*. 1. ed. [S.l.]: Springer, 2009. Hardcover.

DIACONIS, P.; ZABELL, S. L. Updating subjective probability. *Journal of the American Statistical Association*, Taylor & Francis, v. 77, p. 822–830, 1982.

DORE, L. H. et al. Bias-corrected maximum likelihood estimation of the parameters of the complex Bingham distribution. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 30, p. 385–400, 2016.

DRYDEN, I. Discussion to 'Procrustes methods in the statistical analysis of shape' by CR Goodall. *Journal of the Royal Statistical Society. Series B (Methodological),*, v. 53, p. 327–328, 1991.

DRYDEN, I.; MARDIA, K. V. General shape distributions in a plane. *Advances in Applied Probability*, Cambridge University Press, v. 23, p. 259–276, 1991.

DRYDEN, I. L. et al. Mean shapes, projections and intrinsic limiting distributions. *Journal of Statistical Planning and Inference*, p. 25–32, 2014.

DRYDEN, I. L.; MARDIA, K. V. *Statistical Shape Analysis*. [S.l.]: J. Wiley Chichester, 1998.

DRYDEN, I. L.; MARDIA, K. V. *Statistical Shape Analysis: With Applications in R*. [S.l.]: John Wiley & Sons, 2016.

DRYDEN, I. L. et al. Statistical analysis on high-dimensional spheres and shape spaces. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 33, p. 1643–1665, 2005.

EGUCHI, S.; COPAS, J. Interpreting Kullback–Leibler divergence with the Neyman–Pearson lemma. *Journal of Multivariate Analysis*, Academic Press, v. 97, p. 2034–2040, 2006.

EMMERT-STREIB, F.; DEHMER, M. *Information Theory and Statistical Learning*. [S.l.]: Springer, 2009.

ESTEBAN, M. On testing hypotheses with divergence statistics. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 25, p. 501–518, 1996.

FALCONER, D. Replicated selection for body weight in mice. *Genetics Research*, Cambridge University Press, v. 22, p. 291–321, 1973.

FISHER, N. I. et al. Improved Pivotal Methods for Constructing Confidence Regions with Directional Data. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 91, p. 1062–1070, 1996.

FRANGI, A. F. et al. Automatic Construction of Multiple-Object Three-Dimensional Statistical Shape Models: Application to Cardiac Modeling. *IEEE Transactions on Medical Imaging*, Citeseer, v. 21, p. 1151–1166, 2002.

FRERY, A. C.; CINTRA, R. J.; NASCIMENTO, A. D. Entropy-Based Statistical Analysis of PolSAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, v. 51, p. 3733–3743, 2013.

FRERY, A. C.; NASCIMENTO, A. D.; CINTRA, R. J. Analytic Expressions for Stochastic Distances Between Relaxed Complex Wishart Distributions. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, v. 52, p. 1213–1226, 2014.

GOLLAND, P. et al. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, Elsevier, v. 9, p. 69–86, 2005.

GOLSHANI, L.; PASHA, E. Rényi entropy rate for Gaussian processes. *Information Sciences*, Elsevier, v. 180, p. 1486–1491, 2010.

GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. [S.l.]: JHU Press, 2012.

GOODALL, C. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, v. 53, p. 285–339, 1991.

GRÄTZER, G. *More Math into Latex*. [S.l.]: Springer Science & Business Media, 2007.

GUL, G.; ZOUBIR, A. M. Robust Hypothesis Testing with α-Divergence. *IEEE Transactions on Signal Processing*, Institute of Electrical and Electronics Engineers (IEEE), v. 64, p. 4737–4750, 2016.

HARTLEY, R. V. L. Transmission of Information. *Bell System Technical Journal*, v. 7, p. 535–563, 1928.

HURLEY, J. R.; CATTELL, R. B. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, Wiley Online Library, v. 7, p. 258–262, 1962.

JAMES, G. Tests of Linear Hypotheses in Univariate and Multivariate Analysis when the Ratios of the Population Variances are Unknown. *Biometrika*, JSTOR, v. 41, p. 19–43, 1954.

JOHNSON, D. et al. Measurement of biological shape: A general method applied to mouse vertebrae. *Development*, The Company of Biologists Ltd, v. 90, p. 363–377, 1985.

JOHNSON, D.; O'HIGGINS, P.; MCANDREW, T. The effect of replicated selection for body weight in mice on vertebral shape. *Genetics Research*, Cambridge University Press, v. 51, p. 129–135, 1988.

KAILATH, T. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Transactions on Communication Technology*, IEEE, v. 15, p. 52–60, 1967.

KENDALL, D. G. The shape of Poisson-Delaunay triangles. *Studies in Probability and Related Topics (eds M. C. Demetreseu and M. Iosifescu)*, Nagard, Montreal., v. 23, p. 321–330, 1983.

KENDALL, D. G. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, Wiley Online Library, v. 16, p. 81–121, 1984.

KENT, J. T. Asymptotic Expansions for the Bingham distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 36, p. 139–144, 1987.

KENT, J. T. The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, v. 56, p. 285–299, 1994.

KENT, J. T. Current Issues for Statistical Inference in Shape Analysis. *Proceedings in Current Issues in Statistical Shape Analysis*, Leeds University Press Leeds, p. 167–175, 1995.

KENT, J. T. Data analysis for shapes and images. *Journal of statistical planning and inference*, Elsevier, v. 57, n. 2, p. 181–193, 1997.

KENT, J. T.; CONSTABLE, P. D.; ER, F. Simulation for the complex Bingham distribution. *Statistics and Computing*, Springer, v. 14, p. 53–57, 2004.

KULLBACK, S. *Information Theory and Statistics*. [S.l.]: Courier Corporation, 1978.

KUME, A.; WOOD, A. T. Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. *Biometrika*, Biometrika Trust, v. 92, p. 465–476, 2005.

KUME, A.; WOOD, A. T. On the derivatives of the normalising constant of the Bingham distribution. *Statistics and Probability Letters*, Elsevier, v. 77, p. 832–837, 2007.

KURZ, G.; HANEBECK, U. D. Stochastic Sampling of the Hyperspherical von Mises–Fisher Distribution Without Rejection Methods. In: IEEE. *Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015*. [S.l.], 2015. p. 1–6.

KURZ, G.; PFAFF, F.; HANEBECK, U. D. Kullback-Leibler Divergence and Moment Matching for Hyperspherical Probability Distributions. In: IEEE. *Information Fusion (FUSION), 2016 19th International Conference on*. [S.l.], 2016. p. 2087–2094.

LE, H.; KUME, A. The fréchet mean shape and the shape of the means. *Advances in Applied Probability*, Cambridge University Press, v. 32, p. 101–113, 2000.

LELE, S.; III, T. M. C. A new test for shape differences when variance-covariance matrices are unequal. *Journal of Human Evolution*, Elsevier, v. 31, p. 193–212, 1996.

LELE, S.; RICHTSMEIER, J. T. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, Wiley Online Library, v. 86, p. 415–427, 1991.

LEU, R.; DAMIEN, P. Bayesian shape analysis of the complex Bingham distribution. *Journal of Statistical Planning and Inference*, Elsevier, v. 149, p. 183–200, 2014.

LIESE, F.; VAJDA, I. On Divergences and Informations in Statistics and Information Theory. *IEEE Transactions on Information Theory*, IEEE, v. 52, p. 4394–4412, 2006.

LIN, J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, IEEE, v. 37, p. 145–151, 1991.

LOHMANN, G. Eigenshape analysis of microfossils: A general morphometric procedure for describing changes in shape. *Mathematical Geology*, Springer, v. 15, p. 659–672, 1983.

MACKAY, D. J. *Information Theory, Inference and Learning Algorithms*. [S.l.]: Cambridge University Press, 2003.

MARDIA, K. Directional statistics and shape analysis. *Journal of Applied Statistics*, Taylor & Francis, v. 26, p. 949–957, 1999.

MARDIA, K.; DRYDEN, I. Shape distributions for landmark data. *Advances in Applied Probability*, Cambridge University Press, v. 21, p. 742–755, 1989.

MARDIA, K.; DRYDEN, I. The statistical analysis of shape data. *Biometrika*, Oxford University Press, v. 76, p. 271–281, 1989.

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, p. 913–926, 1999.

MARDIA, K. V. Shape Analysis of Triangles Through Directional Techniques. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 51, p. 449–458, 1989.

MENÉNDEZ, M. L. et al. Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: a unified study. *Statistical Papers*, v. 36, p. 1–29, 1995.

MICHEAS, A. C.; DEY, D. K. Assessing Shape Differences in Populations of Shapes Using the Complex Watson Shape Distribution. *Journal of Applied Statistics*, Taylor & Francis, v. 32, p. 105–116, 2005.

MICHEAS, A. C.; DEY, D. K. Modeling shape distributions and inferences for assessing differences in shapes. *Journal of Multivariate Analysis*, Elsevier, v. 92, p. 257–280, 2005.

MICHEAS, A. C.; DEY, D. K.; MARDIA, K. V. Complex elliptical distributions with application to shape analysis. *Journal of Statistical Planning and Inference*, Elsevier, v. 136, p. 2961–2982, 2006.

MIDDLETON, D.; ELECTRICAL, I. of; ENGINEERS, E. *An Introduction to Statistical Communication Theory*. [S.l.]: IEEE press Piscataway, NJ, 1996.

MORALES, D.; PARDO, L.; PARDO, M. Likelihood divergence statistics for testing hypotheses about multiple population. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 30, p. 867–884, 2001.

NADARAJAH, S.; ZOGRAFOS, K. Expressions for Rényi and Shannon entropies for bivariate distributions. *Information Sciences*, Elsevier, v. 170, p. 173–189, 2005.

NASCIMENTO, A. D.; CINTRA, R. J.; FRERY, A. C. Hypothesis Testing in Speckled Data With Stochastic Distances. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, v. 48, p. 373–385, 2010.

NOCEDAL, J.; WRIGHT, S. *Numerical optimization*. [S.l.]: Springer Science & Business Media, 2006.

O'HIGGINS, P. *A morphometric study of cranial shape in the Hominoidea*. University of Leeds, 1989.

O'HIGGINS, P.; DRYDEN, I. L. Sexual dimorphism in hominoids: further studies of craniofacial shape differences in Pan, Gorilla and Pongo. *Journal of Human Evolution*, Elsevier, v. 24, p. 183–205, 1993.

O'HIGGINS, P. et al. Patterns of cranial sexual dimorphism in certain groups of extant hominoids. *Journal of Zoology*, Wiley Online Library, v. 222, p. 399–420, 1990.

OXNARD, C. E. Sexual dimorphisms in the overall proportions of primates. *American Journal of Primatology*, Wiley Online Library, v. 4, p. 1–22, 1983.

PARDO, L. *Statistical Inference based on Divergence Measures*. [S.l.]: CRC Press, 2005.

PIETRZAK, M. et al. Limit theorems for empirical Rényi entropy and divergence with applications to molecular diversity analysis. *Test*, Springer, v. 25, p. 654–673, 2016.

RÉNYI, A. et al. On Measures of Entropy and Information. In: *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.: s.n.], 1961. p. 547–561.

ROHLF, F. J. Shape Statistics: Procrustes Superimpositions and Tangent Spaces. *Journal of Classification*, Springer, v. 16, p. 197–223, 1999.

ROHLF, F. J. On the use of shape spaces to compare morphometric methods. *Hystrix-the Italian Journal of Mammalogy*, v. 11, 2000.

ROHLF, F. J. Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, Wiley Online Library, v. 111, p. 463–478, 2000.

ROHLF, F. J. Bias and error in estimates of mean shape in geometric morphometrics. *Journal of Human Evolution*, Elsevier, v. 44, p. 665–683, 2003.

RÖVER, C.; FRIEDE, T. Discrete Approximation of a Mixture Distribution via Restricted Divergence. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 26, p. 217–222, 2017.

SALICRU, M. et al. Asymptotic distribution of (h, φ)-entropies. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 22, p. 2015–2031, 1993.

SALICRÚ, M. et al. On the Applications of Divergence Type Measures in Testing Statistical Hypotheses. *Journal of Multivariate Analysis*, Elsevier, v. 51, p. 372–391, 1994.

SÁNCHEZ-MORENO, P.; ANGULO, J. C.; DEHESA, J. S. A generalized complexity measure based on Rényi entropy. *The European Physical Journal D*, Springer, v. 68, p. 212, 2014.

SEBER, G. A. *Multivariate Observations*. [S.l.]: John Wiley & Sons, 1984.

SEBER, G. A. *A matrix handbook for statisticians*. [S.l.]: John Wiley & Sons, 2008.

SEGHOUANE, A.-K.; AMARI, S.-I. The AIC criterion and symmetrizing the Kullback–Leibler divergence. *IEEE Transactions on Neural Networks*, IEEE, v. 18, p. 97–106, 2007.

SEO, J.-I.; KANG, S.-B. Entropy Estimation of Generalized Half-Logistic Distribution (GHLD) Based on Type-II Censored Samples. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 16, p. 443–454, 2014.

SHANNON, C. E. A Mathematical Theory of Communication, Part I, Part II. *The Bell System Technical Journal*, v. 27, p. 623–656, 1948.

SONAT, F. A. et al. Statistical shape analysis of the rat hippocampus in epilepsy. *Anatomical Science International*, Springer, v. 84, p. 298, 2009.

SONG, S.; SONG, X.; KANG, Y. Entropy-Based Parameter Estimation for the Four-Parameter Exponential Gamma Distribution. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 19, p. 189, 2017.

SYROPOULOS, A.; TSOLOMITIS, A.; SOFRONIOU, N. *Digital typography using LaTeX*. [S.l.]: Springer Science & Business Media, 2003.

TANEJA, I. J. Generalized Symmetric Divergence Measures and the Probability of Error. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 42, p. 1654–1672, 2013.

TOMA, A.; BRONIATOWSKI, M. Dual divergence estimators and tests: Robustness results. *Journal of Multivariate Analysis*, Elsevier BV, v. 102, p. 20–36, 2011.

WANG, L. et al. Automatic Gait Recognition Based on Statistical Shape Analysis. *IEEE Transactions on Image Processing*, v. 12, p. 1120–1131, 2003.

WATSON, G. Equatorial distributions on a sphere. *Biometrika*, JSTOR, v. 52, p. 193–201, 1965.

WATSON, G. The shapes of a random sequence of triangles. *Advances in Applied Probability*, Cambridge University Press, v. 18, p. 156–169, 1986.

APPENDIX A – PROOF OF THE RESULTS OF NEW DISTANCES ON $\mathbb{C}W$ DISTRIBUTION

Proof of the Theorem 3.1.

The Kullback-Leibler divergence between two elements in $\boldsymbol{z} \sim \mathbb{C}W$ having densities $f(\boldsymbol{z};\boldsymbol{\theta}_1)$ and $f(\boldsymbol{z};\boldsymbol{\theta}_2)$ and using the definition 2.8 we get

$$\begin{aligned} D_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) &= \mathbb{E}_{\boldsymbol{\theta}_1}\left[\log\left(\frac{f(\boldsymbol{z};\boldsymbol{\theta}_1)}{f(\boldsymbol{z};\boldsymbol{\theta}_2)}\right)\right] = \log\left(\frac{c_1(\kappa_2)}{c_1(\kappa_1)}\right) + \mathbb{E}_{\boldsymbol{\theta}_1}\left[\kappa_1|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 - \kappa_2|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2\right] \\ &= \log\left(\frac{c_1(\kappa_2)}{c_1(\kappa_1)}\right) + \kappa_1\mathbb{E}_{\boldsymbol{\theta}_1}[|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2] - \kappa_2\mathbb{E}_{\boldsymbol{\theta}_1}[|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2]. \end{aligned}$$

Analogously,

$$D_{KL}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) = \log\left(\frac{c_1(\kappa_1)}{c_1(\kappa_2)}\right) + \kappa_2\mathbb{E}_{\boldsymbol{\theta}_2}(|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2) - \kappa_1\mathbb{E}_{\boldsymbol{\theta}_2}(|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)$$

Thus

$$\begin{aligned} d_{KL}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= \frac{1}{2}[D_{KL}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) + D_{KL}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)] \\ &= \frac{1}{2}\left\{\kappa_1\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}] - \kappa_2\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}] + \kappa_2\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}) - \kappa_1\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z})\right\} \end{aligned} \tag{A.1}$$

In order to find expected values in (A.1), we rewrite their arguments in terms of quadratic forms and use the Kent's polar shape coordinates (for details see Kent (1994), Dryden and Mardia (2016)). Given $\boldsymbol{z} = [z_1,\ldots,z_k]^\top \sim \mathbb{C}W$, set $s_j = z_j^* z_j$ for $j = 1,\ldots,k-1$. The vector $\boldsymbol{s} = [s_1,\ldots,s_{k-1}]^\top$ follows the joint TME to a unit simplex Dryden and Mardia (2016), denoted as $\mathcal{S}_{k-1}$ and defined by

$$\mathcal{S}_{k-1} = \{\boldsymbol{s} = [s_1,\ldots,s_{k-1}]^\top : s_j \geq 0 \text{ and } \sum_{j=1}^{k-1} s_j \leq 1\}.$$

So, the following quadratic form holds

$$\boldsymbol{z}^*\boldsymbol{T}\boldsymbol{z} = \sum_{j=1}^{k}\lambda_j z_j^* z_j = \sum_{j=1}^{k}\lambda_j s_j.$$

This implies that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}) &= \mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{T}_1\boldsymbol{z}) \\
&= \int_{\mathbb{C}S^{k-1}} (\boldsymbol{z}^*\boldsymbol{T}_1\boldsymbol{z}) c_1^{-1}(\kappa_2) \exp(\kappa_2 |\boldsymbol{z}\boldsymbol{\mu}_2^*|^2) w_k(d\boldsymbol{z}) \\
&= c_1^{-1}(\kappa_2) \int_{\mathbb{C}S^{k-1}} (\boldsymbol{z}^*\boldsymbol{T}_1\boldsymbol{z}) \exp(\boldsymbol{z}^*\kappa_2\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}) w_k(d\boldsymbol{z}) \\
&= c_1^{-1}(\kappa_2) \int_{\mathbb{C}S^{k-1}} (\boldsymbol{z}^*\boldsymbol{T}_1\boldsymbol{z}) \exp(\boldsymbol{z}^*\boldsymbol{T}_2\boldsymbol{z}) w_k(d\boldsymbol{z}) \\
&= c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{T_{1j}} \int_{\mathcal{S}_{k-1}\times[0,2\pi)^k]} s_j \exp(\sum_{j=1}^{k} \lambda_{T_{2j}} s_j) 2^{1-k} \nu_{k-1}(d\boldsymbol{s}) d\boldsymbol{\theta} \\
&= c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{T_{1j}} 2\pi^k \int_{\mathcal{S}_{k-1}} s_j \exp(\sum_{j=1}^{k} \lambda_{T_{2j}} s_j) \nu_{k-1}(d\boldsymbol{s}) \\
&= c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{T_{1j}} 2\pi^k \int_{\mathcal{S}_{k-1}} \frac{\partial}{\partial \lambda_{T_{2j}}} \left[\exp(\sum_{j=1}^{k} \lambda_{T_{2j}} s_j) \right] \nu_{k-1}(d\boldsymbol{s}) \\
&= c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{T_{1j}} \frac{\partial}{\partial \lambda_{T_{2j}}} \left[2\pi^k \int_{\mathcal{S}_{k-1}} \exp(\sum_{j=1}^{k} \lambda_{T_{2j}} s_j) \nu_{k-1}(d\boldsymbol{s}) \right] \\
&= c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{T_{1j}} \frac{\partial}{\partial \lambda_{T_{2j}}} c(\boldsymbol{\Lambda}_{T_2}),
\end{aligned}$$

where $\nu_{k-1}(d\boldsymbol{s})$ is the uniform measure on $\boldsymbol{\mathcal{S}}_{k-1}$, $\boldsymbol{T}_1 = \boldsymbol{\mu}_1\boldsymbol{\mu}_1^*$, $\boldsymbol{T}_2 = \kappa_2\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*$ and - assuming $\boldsymbol{\Lambda}_B = \text{diag}(\lambda_{B_1}, \ldots, \lambda_{B_k})$ as the diagonal matrix of eigenvalues of $\boldsymbol{B}$ – the first derivatives of $c(\boldsymbol{\Lambda}_B)$ are given by Amaral, Floréz and Cysneiros (2013)

$$\frac{\partial c(\boldsymbol{\Lambda}_B)}{\partial \lambda_{B_i}} = 2\pi^k \left[\sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{a_j \text{exp}(\lambda_{B_j})}{(\lambda_{B_j} - \lambda_{B_i})} + a_i \text{exp}(\lambda_{B_i}) + a_i b_i \exp(\lambda_{B_i}) \right],$$

with $a_j^{-1} = \prod_{\substack{l=1 \\ l\neq j}}^{k} (\lambda_{B_j} - \lambda_{B_l})$ and $b_i = \sum_{l\neq i} \frac{1}{(\lambda_{B_i} - \lambda_{B_l})}$. Using the same way, we have

$$\mathbb{E}_{\boldsymbol{\theta}_1}(\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}) = c_1^{-1}(\kappa_1) \sum_{j=1}^{k} \lambda_{T_{1j}} \frac{\partial}{\partial \lambda_{\boldsymbol{M}_{1j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{M}_1})$$

for $\boldsymbol{M}_1 = \kappa_1\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*$,

$$\mathbb{E}_{\boldsymbol{\theta}_1}(\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}) = c_1^{-1}(\kappa_1) \sum_{j=1}^{k} \lambda_{\boldsymbol{M}_{2j}} \frac{\partial}{\partial \lambda_{\boldsymbol{M}_{1j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{M}_1})$$

for $\boldsymbol{M}_2 = \boldsymbol{\mu}_2\boldsymbol{\mu}_2^*$ and

$$\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}) = c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{\boldsymbol{M}_{2j}} \frac{\partial}{\partial \lambda_{\boldsymbol{T}_{2j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{T}_2})$$

for $\boldsymbol{M}_2 = \boldsymbol{\mu}_2\boldsymbol{\mu}_2^*$. Therefore, the Kullback-Leibler distance is

$$\begin{aligned} d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) =& \frac{1}{2}\left\{\kappa_1 \mathbb{E}_{\boldsymbol{\theta}_1}[|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2] - \kappa_2 \mathbb{E}_{\boldsymbol{\theta}_1}[|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2] + \kappa_2 \mathbb{E}_{\boldsymbol{\theta}_2}(|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2) - \kappa_1 \mathbb{E}_{\boldsymbol{\theta}_2}(|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)\right\} \\ =& \frac{1}{2}\left\{\kappa_1 c_1^{-1}(\kappa_1) \sum_{j=1}^{k} \lambda_{\boldsymbol{T}_{1j}} \frac{\partial}{\partial \lambda_{\boldsymbol{M}_{1j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{M}_1}) - \kappa_2 c_1^{-1}(\kappa_1) \sum_{j=1}^{k} \lambda_{\boldsymbol{M}_{2j}} \frac{\partial}{\partial \lambda_{\boldsymbol{M}_{1j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{M}_1}) \right. \\ & \left. + \kappa_2 c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{\boldsymbol{M}_{2j}} \frac{\partial}{\partial \lambda_{\boldsymbol{T}_{2j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{T}_2}) - \kappa_1 c_1^{-1}(\kappa_2) \sum_{j=1}^{k} \lambda_{\boldsymbol{T}_{1j}} \frac{\partial}{\partial \lambda_{\boldsymbol{T}_{2j}}} c(\boldsymbol{\Lambda}_{\boldsymbol{T}_2}) \right\}. \end{aligned}$$

■

Proof of the Theorem 3.2.

The the Rényi divergence with order $\beta \in (0,1)$ is given by

$$\begin{aligned} D_R^{\beta}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) &= \frac{1}{\beta-1} \log \mathbb{E}_{\boldsymbol{\theta}_1}\left\{\left[\frac{f(\boldsymbol{z};\boldsymbol{\theta}_1)}{f(\boldsymbol{z};\boldsymbol{\theta}_2)}\right]^{\beta-1}\right\} \\ &= \frac{1}{\beta-1} \log \mathbb{E}_{\boldsymbol{\theta}_1}\left\{\left[\frac{c_1^{-1}(\kappa_1)\exp(\kappa_1|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)}{c_1^{-1}(\kappa_2)\exp(\kappa_2|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)}\right]^{\beta-1}\right\} \\ &= \frac{1}{\beta-1} \log \left[\left(\frac{c_1^{-1}(\kappa_1)}{c_1^{-1}(\kappa_2)}\right)^{\beta-1} \cdot \mathbb{E}_{\boldsymbol{\theta}_1}[\exp(\kappa_1(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 - \kappa_2(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)\right], \end{aligned} \tag{A.2}$$

where

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\theta}_1}[\exp(\kappa_1(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 - \kappa_2(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)] \\ &= \int_{\mathbb{C}S^{k-1}} \exp(\kappa_1(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 - \kappa_2(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2) c_1^{-1}(\kappa_1)\exp(\kappa_1|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)\ w_k(d\boldsymbol{z}) \\ &= c_1^{-1}(\kappa_1) \int_{\mathbb{C}S^{k-1}} \exp(\kappa_1\beta|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 - \kappa_2(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)\ w_k(d\boldsymbol{z}) \\ &= c_1^{-1}(\kappa_1) \int_{\mathbb{C}S^{k-1}} \exp[\boldsymbol{z}^*(\kappa_1\beta\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* - \kappa_2(\beta-1)\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*)\boldsymbol{z}]\ w_k(d\boldsymbol{z}) \\ &= c_1^{-1}(\kappa_1)c(\boldsymbol{M}_1) \int_{\mathbb{C}S^{k-1}} c^{-1}(\boldsymbol{M}_1)\exp(\boldsymbol{z}^*\boldsymbol{M}_1\boldsymbol{z})\ w_k(d\boldsymbol{z}) \\ &= c_1^{-1}(\kappa_1)c(\boldsymbol{M}_1), \end{aligned} \tag{A.3}$$

where $\boldsymbol{M}_1 = (\kappa_1\beta\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* - \kappa_2(\beta-1)\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*)$ is a hermitian matrix of order $k \times k$. From (A.2) and (A.3), we have

$$D_R^{\beta}(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) = \frac{1}{\beta-1} \log \left[\left(\frac{c_1^{-1}(\kappa_1)}{c_1^{-1}(\kappa_2)}\right)^{\beta-1} c_1^{-1}(\kappa_1)c(\boldsymbol{M}_1)\right] = \frac{1}{\beta-1} \log \left(\frac{c_1^{\beta-1}(\kappa_2)c(\boldsymbol{M}_1)}{c_1^{\beta}(\kappa_1)}\right).$$

Analogously,

$$D_R^{\beta}(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) = \frac{1}{\beta-1} \log \left(\frac{c_1^{\beta-1}(\kappa_1)c(\boldsymbol{M}_2)}{c_1^{\beta}(\kappa_2)}\right),$$

where $\boldsymbol{M}_2 = (\kappa_2\beta\boldsymbol{\mu}_2\boldsymbol{\mu}_2^* - \kappa_1(\beta-1)\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*)$. Therefore, the Rényi distance is given by

$$\begin{aligned} d_R^\beta(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= \frac{1}{\beta-1}\log\left\{\frac{\exp[(\beta-1)D_R^\beta(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)] + \exp[(\beta-1)D_R^\beta(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)]}{2}\right\} \\ &= \frac{1}{\beta-1}\log\left[\frac{1}{2}\left(\frac{c_1^{\beta-1}(\kappa_2)\,c(\boldsymbol{M}_1)}{c_1^\beta(\kappa_1)} + \frac{c_1^{\beta-1}(\kappa_1)\,c(\boldsymbol{M}_2)}{c_1^\beta(\kappa_2)}\right)\right]. \end{aligned}$$

■

Proof of the Corollary 3.1.

The Bhattacharyya distance is

$$\begin{aligned} d_B(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= -\log\left[\int_{\mathbb{C}S^{k-1}}\sqrt{f(\boldsymbol{z};\boldsymbol{\theta}_1)(\boldsymbol{z};\boldsymbol{\theta}_2)}\;w_k(d\boldsymbol{z})\right] \\ &= -\log\left[\int_{\mathbb{C}S^{k-1}}\sqrt{c_1^{-1}(\kappa_1)\exp(\kappa_1|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)c_1^{-1}(\kappa_2)\exp(\kappa_2|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)}\;w_k(d\boldsymbol{z})\right] \\ &= -\log\left[c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)\int_{\mathbb{C}S^{k-1}}\exp\left(\frac{\kappa_1}{2}|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 + \frac{\kappa_2}{2}|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2\right)\;w_k(d\boldsymbol{z})\right] \\ &= -\log\left\{c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)\int_{\mathbb{C}S^{k-1}}\exp\left[\boldsymbol{z}^*\left(\frac{\kappa_1}{2}\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* + \frac{\kappa_2}{2}\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\right)\boldsymbol{z}\right]\;w_k(d\boldsymbol{z})\right\} \\ &= -\log\left\{c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)c(\boldsymbol{A})\int_{\mathbb{C}S^{k-1}}c^{-1}(\boldsymbol{A})\exp\left(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}\right)\;w_k(d\boldsymbol{z})\right\} \\ &= -\log\left(\frac{c(\boldsymbol{A})}{\sqrt{c_1(\kappa_1)c_1(\kappa_2)}}\right), \end{aligned}$$

where $\boldsymbol{A} = \left(\frac{\kappa_1}{2}\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* + \frac{\kappa_2}{2}\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\right)$.
The Hellinger distance is given by

$$\begin{aligned} d_H(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= 1 - \int\limits_{\mathbb{C}S^{k-1}}\sqrt{f(\boldsymbol{z};\boldsymbol{\theta}_1)f(\boldsymbol{z};\boldsymbol{\theta}_2)}\;w_k(d\boldsymbol{z}) \\ &= 1 - \int\limits_{\mathbb{C}S^{k-1}}\sqrt{c_1^{-1}(\kappa_1)\exp(\kappa_1|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2)c_1^{-1}(\kappa_2)\exp(\kappa_2|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2)}\;w_k(d\boldsymbol{z}) \\ &= 1 - c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)\int\limits_{\mathbb{C}S^{k-1}}\exp\left(\frac{\kappa_1}{2}|\boldsymbol{z}^*\boldsymbol{\mu}_1|^2 + \frac{\kappa_2}{2}|\boldsymbol{z}^*\boldsymbol{\mu}_2|^2\right)\;w_k(d\boldsymbol{z}) \\ &= 1 - c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)\int\limits_{\mathbb{C}S^{k-1}}\exp\left[\boldsymbol{z}^*\left(\frac{\kappa_1}{2}\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* + \frac{\kappa_2}{2}\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\right)\boldsymbol{z}\right]\;w_k(d\boldsymbol{z}) \\ &= 1 - c_1^{-1/2}(\kappa_1)c_1^{-1/2}(\kappa_2)c(\boldsymbol{A})\int\limits_{\mathbb{C}S^{k-1}}c^{-1}(\boldsymbol{A})\exp(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z})\;w_k(d\boldsymbol{z}), \end{aligned}$$

where $\boldsymbol{A} = \left(\frac{\kappa_1}{2}\boldsymbol{\mu}_1\boldsymbol{\mu}_1^* + \frac{\kappa_2}{2}\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\right)$. So

$$d_H(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = 1 - \frac{c(\boldsymbol{A})}{\sqrt{(c_1(\kappa_1)c_1(\kappa_2))}}.$$

■

Proof of the Proposition 3.1. Let $\boldsymbol{R}$ be a rotation matrix,

$$
\begin{aligned}
d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}) &= \frac{1}{2}\{\kappa_1\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}] - \kappa_2\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}] \\
&+ \kappa_2\mathbb{E}_{\boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}}(\boldsymbol{z}^*\boldsymbol{R}^*(\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*)\boldsymbol{R}\boldsymbol{z}) - \kappa_1\mathbb{E}_{\boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}}(\boldsymbol{z}^*\boldsymbol{R}^*(\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*)\boldsymbol{R}\boldsymbol{z})\} \\
&= \frac{1}{2}\{\kappa_1\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}] - \kappa_2\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}] \\
&+ \kappa_2\mathbb{E}_{\boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}}((\boldsymbol{R}\boldsymbol{z})^*(\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*)\boldsymbol{R}\boldsymbol{z}) - \kappa_1\mathbb{E}_{\boldsymbol{R}^*\boldsymbol{\theta}_2\boldsymbol{R}}((\boldsymbol{R}\boldsymbol{z})^*(\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*)\boldsymbol{R}\boldsymbol{z})\} \\
&= \frac{1}{2}\{\kappa_1\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z}] - \kappa_2\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}] + \kappa_2\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_2\boldsymbol{\mu}_2^*\boldsymbol{z}) - \kappa_1\mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{z}^*\boldsymbol{\mu}_1\boldsymbol{\mu}_1^*\boldsymbol{z})\} \\
&= d_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).
\end{aligned}
$$

■

Proof of the Proposition 2.2. Let $\boldsymbol{A} = \kappa(\boldsymbol{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)$ be a matrix with order two and parameter of the $\mathbb{C}B$ distribution and $\boldsymbol{\mu} = [a + \mathrm{i}b, c + \mathrm{i}d]^\top$.

Consider the problem of determining the eigenvalue of $\boldsymbol{A}$; i.e., finding the root of $\det(\boldsymbol{A} - \lambda_1\boldsymbol{I}) = 0$.

After some algebraic manipulations, we arrive in the second degree equation

$$\lambda_1^2 - \lambda_1(x + y) + xy - \kappa^2||u||^2 = 0, \tag{A.4}$$

where

$$\begin{cases} x = \kappa - \kappa(a^2 + b^2), \\ y = \kappa - \kappa(c^2 + d^2), \\ u = ac + bd + \mathrm{i}(bc - cd). \end{cases}$$

Solving the equation (A.4) and taking the positive root, we find the proportional relation between λ_1 and κ:

$$\kappa = \frac{2\lambda_1}{g(\boldsymbol{\mu})},$$

where $g(\mu) = 1 + \sqrt{(c^2 + d^2 - a^2 - b^2)^2 + 4||u||^2}$. where $g(\mu) = [2 - (a^2 + b^2 + c^2 + d^2)] + \sqrt{(c^2 + d^2 - a^2 - b^2)^2 + 4||u||^2}$. ■

APPENDIX B – PROOF OF THE RESULTS OF NEW DISTANCES ON $\mathbb{C}B$ DISTRIBUTION

Proof of the Theorem 5.1. Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_2)$ and their densities given by $f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)$, respectively, having common support $\boldsymbol{z} = [z_1, z_2, \ldots, z_k]^\top \in \mathbb{C}\boldsymbol{S}^{k-1}$. The KL divergence measures are given by

$$\begin{aligned} D_{KL}(\boldsymbol{A}_1||\boldsymbol{A}_2) &= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1) \log\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)}\right) w_k(d\boldsymbol{z}) \\ &= \mathbb{E}_{\boldsymbol{A}_1}\left[\log\left(\frac{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)}{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)}\right)\right] \\ &= \mathbb{E}_{\boldsymbol{A}_1}\left\{\log[\, c^{-1}(\boldsymbol{A}_1)\exp(\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z})\,] - \log[\, c^{-1}(\boldsymbol{A}_2)\exp(\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z})\,]\right\} \\ &= \log c(\boldsymbol{A}_2) - \log c(\boldsymbol{A}_1) + \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}\,). \end{aligned} \tag{B.1}$$

and, analogously,

$$\begin{aligned} D_{KL}(\boldsymbol{A}_2||\boldsymbol{A}_1) &= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2) \log\left(\frac{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)}{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)}\right) w_k(d\boldsymbol{z}) \\ &= \mathbb{E}_{\boldsymbol{A}_2}\left[\log\left(\frac{f_{\boldsymbol{x}_2}(\boldsymbol{z};\boldsymbol{A}_2)}{f_{\boldsymbol{x}_1}(\boldsymbol{z};\boldsymbol{A}_1)}\right)\right] \\ &= \log c(\boldsymbol{A}_1) - \log c(\boldsymbol{A}_2) + \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,). \end{aligned} \tag{B.2}$$

Replacing (B.1) and (B.2) in

$$d_{KL}(\boldsymbol{A}_1, \boldsymbol{A}_2) \;=\; \frac{[\, D_{KL}(\boldsymbol{A}_1||\boldsymbol{A}_2) \;+\; D_{KL}(\boldsymbol{A}_2||\boldsymbol{A}_1)\,]}{2},$$

we have

$$\begin{aligned} d_{KL}(\boldsymbol{A}_1, \boldsymbol{A}_2) &= \frac{1}{2}\,[\log c(\boldsymbol{A}_2) - \log c(\boldsymbol{A}_1) + \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}\,) \\ &\quad + \log c(\boldsymbol{A}_1) - \log c(\boldsymbol{A}_2) + \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,)] \\ &= \frac{1}{2}\,[\mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}\,) + \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,)]\,. \end{aligned} \tag{B.3}$$

In order to derive the expected value at $\boldsymbol{A}_q$ (with $q = 1, 2$) of the quadratic forms, we will use a change of variable by means of the Kent polar coordinates. Now assume

$\boldsymbol{s} = [s_1, \ldots, s_{k-1}]^\top$ follows the TME distribution to a simplex and $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_q)$ for $q = 1, 2$. Further, note the following identities hold:

$$\boldsymbol{z}^* \boldsymbol{A}_q \boldsymbol{z} \;=\; \sum_{j=1}^{k-1} \lambda_j^{(q)} \boldsymbol{z}_j^{(q)^*} \boldsymbol{z}_j^{(q)} = \sum_{j=1}^{k-1} \lambda_j^{(q)}\, s_j^{(q)},$$

where $s_j^{(q)} = ||z_j^{(q)}||^2 = \boldsymbol{z}_j^{(q)^*} \boldsymbol{z}_j^{(q)}$ for $q = 1, 2$ and λ_j are jth eigenvalues of $\boldsymbol{A}$. Thus,

$$\mathbb{E}_{\boldsymbol{A}_q}(\boldsymbol{z}^* \boldsymbol{A}_q \boldsymbol{z}) = \sum_{j=1}^{k-1} \lambda_j^{(q)} \mathbb{E}_{\boldsymbol{\Lambda}_q}(s_j^{(q)}. \tag{B.4}$$

According to Kent (1994) and Dryden and Mardia (2016), we have that

$$\mathbb{E}_{\boldsymbol{\Lambda}_q}(\boldsymbol{s}_j^{(q)}) = \frac{\partial \log c(\boldsymbol{\Lambda}_{(q)})}{\partial \lambda_j^{(q)}} = \frac{1}{c(\boldsymbol{\Lambda}_{(q)})} \frac{\partial c(\boldsymbol{\Lambda}_{(q)})}{\partial \lambda_j^{(q)}} \quad \text{for} \quad q = 1, 2, \tag{B.5}$$

where

$$c(\boldsymbol{\Lambda}) = 2\pi^k \sum_{j=1}^{k} a_j \exp(\lambda_j) \text{ and } \quad a_j^{-1} = \prod_{i \neq j} (\lambda_j - \lambda_i). \tag{B.6}$$

The first derivative of (B.6) is given by

$$\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_j} = 2\pi^k \left[\sum_{l \neq j}^{k} \frac{a_l \exp(\lambda_l)}{(\lambda_l - \lambda_j)} + a_j \exp(\lambda_j) - a_j b_j \exp(\lambda_j) \right],$$

where $b_j = \sum_{i \neq j} (\lambda_j - \lambda_i)^{-1}$. Substituting the expressions (B.5) in (B.4) and posteriorly in (B.3),

$$\begin{aligned} d_{KL}(\boldsymbol{A}_1, \boldsymbol{A}_2) = \frac{1}{2} \Big[\sum_{j=1}^{k} \lambda_j^{(1)} \frac{1}{c(\boldsymbol{\Lambda}_{(1)})} \frac{\partial c(\boldsymbol{\Lambda}_{(1)})}{\partial \lambda_j^{(1)}} - \sum_{j=1}^{k} \lambda_j^{(2)} \frac{1}{c(\boldsymbol{\Lambda}_{(1)})} \frac{\partial c(\boldsymbol{\Lambda}_{(1)})}{\partial \lambda_j^{(1)}} \Big] \\ + \frac{1}{2} \Big[\sum_{j=1}^{k} \lambda_j^{(2)} \frac{1}{c(\boldsymbol{\Lambda}_{(2)})} \frac{\partial c(\boldsymbol{\Lambda}_{(2)})}{\partial \lambda_j^{(2)}} - \sum_{j=1}^{k} \lambda_j^{(1)} \frac{1}{c(\boldsymbol{\Lambda}_{(2)})} \frac{\partial c(\boldsymbol{\Lambda}_{(2)})}{\partial \lambda_j^{(2)}} \Big]. \end{aligned}$$

■

Proof of the Theorem 4.2. Let $\boldsymbol{x}_1 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_1)$ and $\boldsymbol{x}_2 \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A}_2)$ with their density functions given by $f_{\boldsymbol{x}_1}(\boldsymbol{z}; \boldsymbol{A}_1)$ and $f_{\boldsymbol{x}_2}(\boldsymbol{z}; \boldsymbol{A}_2)$, respectively. The Rényi distance with order parameter $\beta \in (0, 1)$ is

$$D_R^\beta(\boldsymbol{A}_1 || \boldsymbol{A}_2) = \frac{1}{(\beta - 1)} \log \int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_1}^\beta(\boldsymbol{z}; \boldsymbol{A}_1) f_{\boldsymbol{x}_2}^{(1-\beta)}(\boldsymbol{z}; \boldsymbol{A}_2)\, w_k(d\boldsymbol{z}),$$

where

$$\begin{aligned} &\int_{\mathbb{C}\boldsymbol{S}^{k-1}} f_{\boldsymbol{x}_1}^\beta(\boldsymbol{z}; \boldsymbol{A}_1) f_{\boldsymbol{x}_2}^{(1-\beta)}(\boldsymbol{z}; \boldsymbol{A}_2)\, w_k(d\boldsymbol{z}) \\ &= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} [c^{-1}(\boldsymbol{A}_1)\{\exp(\boldsymbol{z}^* \boldsymbol{A}_1 \boldsymbol{z})]^\beta [c^{-1}(\boldsymbol{A}_2) \exp(\boldsymbol{z}^* \boldsymbol{A}_2 \boldsymbol{z})]^{(1-\beta)} w_k(d\boldsymbol{z}) \end{aligned}$$

$$
\begin{aligned}
&= c^{-\beta}(\boldsymbol{A}_1)c^{(\beta-1)}(\boldsymbol{A}_2)\int_{\mathbb{C}\boldsymbol{S}^{k-1}}\exp[\boldsymbol{z}^*(\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2)\boldsymbol{z}]\,w_k(d\boldsymbol{z})\\
&= c^{-\beta}(\boldsymbol{A}_1)\,c^{(\beta-1)}(\boldsymbol{A}_2)\,c(\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2)\\
&\cdot\int_{\mathbb{C}\boldsymbol{S}^{k-1}}c^{-1}(\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2)\exp[\boldsymbol{z}^*(\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2)\boldsymbol{z}]\,w_k(d\boldsymbol{z})\\
&= c^{-\beta}(\boldsymbol{A}_1)\,c^{(\beta-1)}(\boldsymbol{A}_2)\,c(\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2)\\
&= c^{-\beta}(\boldsymbol{\Lambda}_1)c^{(\beta-1)}(\boldsymbol{\Lambda}_2)c(\boldsymbol{\Lambda}_{R12}) \qquad \text{(B.7)}
\end{aligned}
$$

where $\boldsymbol{\Lambda}_1$ is the matrix of eigenvalues of $\boldsymbol{A}_1$, $\boldsymbol{\Lambda}_2$ is the matrix of eigenvalues of $\boldsymbol{A}_2$ and $\boldsymbol{\Lambda}_{R12}$ is the matrix of eigenvalues of $\beta\boldsymbol{A}_1+(1-\beta)\boldsymbol{A}_2$.

Analogously,

$$
D_R^{\beta}(\boldsymbol{A}_2||\boldsymbol{A}_1)=\frac{1}{(\beta-1)}\log[c^{-\beta}(\boldsymbol{\Lambda}_2)c^{(\beta-1)}(\boldsymbol{\Lambda}_1)c(\boldsymbol{\Lambda}_{R21})], \qquad \text{(B.8)}
$$

where $\boldsymbol{\Lambda}_{R21}$ is the matrix of eigenvalues of $[\beta\boldsymbol{A}_2+(1-\beta)\boldsymbol{A}_1]$. From (B.7) and (B.8), we obtain the Rényi distance

$$
\begin{aligned}
d_R^{\beta}(\boldsymbol{A}_1,\boldsymbol{A}_2) &= \frac{1}{\beta-1}\log\left\{\frac{\exp\left[(\beta-1)\,D_R^{\beta}(\boldsymbol{A}_1||\boldsymbol{A}_2)\right]\;+\;\exp\left[(\beta-1)\,D_R^{\beta}(\boldsymbol{A}_2||\boldsymbol{A}_1)\right]}{2}\right\}\\
&= \frac{1}{\beta-1}\log\left\{\frac{c^{-\beta}(\boldsymbol{\Lambda}_1)c^{(\beta-1)}(\boldsymbol{\Lambda}_2)c(\boldsymbol{\Lambda}_{R12})\;+\;c^{-\beta}(\boldsymbol{\Lambda}_2)c^{(\beta-1)}(\boldsymbol{\Lambda}_1)c(\boldsymbol{\Lambda}_{R21})}{2}\right\}.
\end{aligned}
$$

■

The proof of Corollary 4.1 follows of taking $\beta=1/2$ and the relations (FRERY; NASCIMENTO; CINTRA, 2014):

$$
d_B(\boldsymbol{A}_1,\boldsymbol{A}_2)=-\log(1-d_H(\boldsymbol{A}_1,\boldsymbol{A}_2))
$$

and

$$
d_H(\boldsymbol{A}_1,\boldsymbol{A}_2)=1-\exp\left(-\frac{1}{2}d_R^{1/2}(\boldsymbol{A}_1,\boldsymbol{A}_2)\right).
$$

To demonstrate the Proposition 4.1, we use the rotation matrix described in Section 3.2 of the Chapter 3.

Proof of the Proposition 4.1. Let $\widehat{\boldsymbol{A}_1}$ and $\widehat{\boldsymbol{A}_2}$ be MLEs for $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, respectively, and be $\boldsymbol{R}$ the rotation matrix . From expression (B.3), we have that

$$
d_{KL}(\boldsymbol{A}_1,\boldsymbol{A}_2)=\frac{1}{2}\left[\mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z})-\mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z}\,)+\mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_2\boldsymbol{z})-\mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,)\right].
$$

Using the property (d) of the Lemma B.1, it follows that

$$\begin{aligned} d_{KL}(\boldsymbol{A}_1, \boldsymbol{A}_2) &= \frac{1}{2}\left[\mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_1}(\,(\boldsymbol{R}\boldsymbol{z})^*\boldsymbol{A}_2\boldsymbol{R}\boldsymbol{z}\,) + \mathbb{E}_{\boldsymbol{A}_2}(\,(\boldsymbol{R}\boldsymbol{z})^*\boldsymbol{A}_2\boldsymbol{R}\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,)\right] \\ &= \frac{1}{2}\left[\mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_1}(\,\boldsymbol{z}^*(\boldsymbol{R}^*\boldsymbol{A}_2\boldsymbol{R})\boldsymbol{z}\,) + \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*(\boldsymbol{R}^*\boldsymbol{A}_2\boldsymbol{R})\boldsymbol{z}) - \mathbb{E}_{\boldsymbol{A}_2}(\,\boldsymbol{z}^*\boldsymbol{A}_1\boldsymbol{z}\,)\right] \\ &= d_{KL}(\boldsymbol{A}_1, \boldsymbol{R}^*\boldsymbol{A}_2\boldsymbol{R}). \end{aligned}$$

■

APPENDIX C – PROOFS RELATED TO THE SHANNON AND RÉNYI ENTROPIES

Proof of the Theorem 5.1. Let $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{B}_{k-1}(\boldsymbol{A})$ be a random variable with pdf $f(\boldsymbol{z}; \boldsymbol{A})$, so the Shannon entropy is given by

$$\begin{aligned} H_S(\boldsymbol{A}) = \mathbb{E}_{\boldsymbol{A}}(-\log f(\boldsymbol{z}; \boldsymbol{A})) &= \mathbb{E}_{\boldsymbol{A}}\{-\log[c^{-1}(\boldsymbol{A})\exp(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z})]\} \\ &= \mathbb{E}_{\boldsymbol{A}}[\log c(\boldsymbol{A}) - \boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}] \\ &= \log c(\boldsymbol{A}) - \mathbb{E}_{\boldsymbol{A}}(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}). \end{aligned}$$

To solve the expectation involving the quadratic forms, we use a variable change based on the Kent's polar shape coordinates. Thus,

$$\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z} = \operatorname{tr}(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}) = \operatorname{tr}\left(\sum_{j=1}^{k} \lambda_j \gamma_j \gamma_j^* z_j^* z_j\right) = \operatorname{tr}\left(\sum_{j=1}^{k} \lambda_j \gamma_j^* \gamma_j z_j^* z_j\right) = \sum_{j=1}^{k} \lambda_j z_j^* z_j = \sum_{j=1}^{k} \lambda_j s_j.$$

Then,

$$\mathbb{E}_{\boldsymbol{A}}(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}) = \sum_{j=1}^{k} \lambda_j \mathbb{E}_{\boldsymbol{\Lambda}}(s_j).$$

According to Kent (1994), we have that

$$\mathbb{E}_{\boldsymbol{\Lambda}}(s_j) = \frac{\partial \log c(\boldsymbol{\Lambda})}{\partial \lambda_j} = \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_j},$$

where the first derivative of normalizing constant is given by

$$\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_j} = 2\pi^k \left[\sum_{l \neq j}^{k} \frac{a_l \exp(\lambda_l)}{(\lambda_l - \lambda_j)} + a_j \exp(\lambda_j) - a_j b_j \exp(\lambda_j)\right].$$

Using the change above transformations, we have to

$$\begin{aligned} H_S(\boldsymbol{A}) = \log c(\boldsymbol{A}) - \mathbb{E}_{\boldsymbol{A}}(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z}) = \log c(\boldsymbol{A}) - \mathbb{E}_{\boldsymbol{\Lambda}}\left(\sum_{i=1}^{k} \lambda_i s_i\right) = \log c(\boldsymbol{A}) - \sum_{i=1}^{k} \lambda_i \mathbb{E}_{\boldsymbol{\Lambda}}(s_i) \\ = \log c(\boldsymbol{A}) - \sum_{i=1}^{k} \lambda_i \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_i} = \log c(\boldsymbol{\Lambda}) - \sum_{i=1}^{k} \lambda_i \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_i}. \end{aligned}$$

Therefore,

$$H_S(\boldsymbol{A}) = \log c(\boldsymbol{\Lambda}) - \sum_{i=1}^{k} \lambda_i \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial \lambda_i}.$$

The Rényi entropy with order parameter $\beta \in \mathbb{R}_+ - \{1\}$ is given by

$$\begin{aligned}
H_R^\beta(\boldsymbol{A}) &= \frac{1}{(1-\beta)} \log\{\mathbb{E}_{\boldsymbol{A}}[f^{\beta-1}(\boldsymbol{z};\boldsymbol{A})]\} = \frac{1}{(1-\beta)} \log\left\{\int_{\mathbb{C}\boldsymbol{S}^{k-1}} [c^{-1}(\boldsymbol{A})\exp(\boldsymbol{z}^*\boldsymbol{A}\boldsymbol{z})]^\beta d\omega_k(\boldsymbol{z})\right\} \\
&= \frac{1}{(1-\beta)} \log\left\{\int_{\mathbb{C}\boldsymbol{S}^{k-1}} c^{-\beta}(\boldsymbol{A})\exp(\boldsymbol{z}^*\beta\boldsymbol{A}\boldsymbol{z}) d\omega_k(\boldsymbol{z})\right\} \\
&= \frac{1}{(1-\beta)} \log\left\{c^{-\beta}(\boldsymbol{A})c(\beta\boldsymbol{A})\int_{\mathbb{C}\boldsymbol{S}^{k-1}} c^{-1}(\beta\boldsymbol{A})\exp(\boldsymbol{z}^*\beta\boldsymbol{A}\boldsymbol{z}) d\omega_k(\boldsymbol{z})\right\} \\
&= \frac{1}{(1-\beta)} \log\frac{c(\beta\boldsymbol{A})}{c^\beta(\boldsymbol{A})} = \frac{1}{(1-\beta)}[\log c(\beta\boldsymbol{A}) - \beta\log c(\boldsymbol{A})],
\end{aligned}$$

where $\omega_k(d\boldsymbol{z})$ is the Lebesgue measure on statistical space $\mathcal{X}, \mathcal{A}$. Therefore, the Rényi entropy is

$$H_R^\beta(\boldsymbol{A}) = \frac{1}{(1-\beta)}[\log c(\beta\boldsymbol{A}) - \beta\log c(\boldsymbol{A})].$$

■

Proof of the Corollary 5.1.

$$\begin{aligned}
\lim_{\beta\to 1} H_R^\beta(\boldsymbol{A}) &\equiv \lim_{\beta\to 1} H_R^\beta(\boldsymbol{\Lambda}) \\
&= \lim_{\beta\to 1} \frac{1}{(1-\beta)}[\log c(\beta\boldsymbol{\Lambda}) - \beta\log c(\boldsymbol{\Lambda})], \quad \text{(applying the L'hospital rule)} \\
&= \lim_{\beta\to 1} \left\{\frac{\partial}{\partial\beta}[\log c(\beta\boldsymbol{\Lambda}) - \beta\log c(\boldsymbol{\Lambda})]\right\} \Big/ \left\{\frac{\partial}{\partial\beta}[(1-\beta)]\right\} \\
&= \lim_{\beta\to 1}\left[\log c(\boldsymbol{\Lambda}) - \frac{1}{c(\beta\boldsymbol{\Lambda})}\frac{\partial c(\beta\boldsymbol{\Lambda})}{\partial\beta}\right] \\
&= \log c(\boldsymbol{\Lambda}) - \lim_{\beta\to 1}\frac{1}{c(\beta\boldsymbol{\Lambda})}\sum_{i=1}^k \lambda_i \frac{\partial c(\beta\boldsymbol{\Lambda})}{\partial\lambda_i} \\
&= \log c(\boldsymbol{\Lambda}) - \sum_{i=1}^k \lambda_i \frac{1}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial\lambda_i} \\
&= H_S(\boldsymbol{\Lambda}) \equiv H_S(\boldsymbol{A}).
\end{aligned}$$

■

The Lebesgue dominated convergence Billingsley (2008) will be used to obtain the Shannon entropy in the Theorem 5.2.

Theorem C.1 (Lebesgue dominated convergence theorem (TCL))**.** *If $|f_n| \leq g$ almost everywhere, where g is integrable, and if $f_n \to f$ almost everywhere, then f and f_n are integrable and $\int f_n \to \int f$.*

Proof of the Theorem 5.2.

(a) Let $\boldsymbol{z} \sim \mathbb{C}\boldsymbol{W}_{k-1}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \kappa)$. The Shannon entropy is obtained as

$$H_S(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\{-\log[c_1^{-1}(\kappa)\exp(\kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2)]\} = \log[c_1(\kappa)] - \kappa\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{z}^*\mu|^2). \tag{C.1}$$

To find a closed-form expression of $\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{z}^*\boldsymbol{\mu}|^2)$, we resort the transform Kent's polar coordinates on $\mathbb{C}\boldsymbol{S}^{k-1}$ given in the proof of the Theorem 5.1 . Initially we take $\boldsymbol{\mu} = [1, 0, \cdots, 0]^T$ and after a suitable rotation $|\boldsymbol{z}^*\boldsymbol{\mu}|^2 = s_1$ Mardia and Dryden (1999, p. 915), this change of variables (in terms of $s_j \in \mathcal{S}_{k-1}$ and $\theta \in [0, 2\pi)$) allows that the complex pre-shape sphere can be identified by $\mathcal{S}_{k-1} \times [0, 2\pi)$ such that the volume measure on $\mathbb{C}\boldsymbol{S}^{k-1}$ is given by

$$2^{(1-k)}ds_1, \ldots ds_{k-1}d\theta_1 \ldots, d\theta_k$$

in this way we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{z}^*\boldsymbol{\mu}|^2) &= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} |\boldsymbol{z}^*\boldsymbol{\mu}|^2 c_1^{-1}(\kappa)\exp(\kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2) w_k(d\boldsymbol{z}) \\
&= \int_{\mathcal{S}_{k-1}\times[0,2\pi)^k} s_1 c_1^{-1}(\kappa)\exp(\kappa s_1) 2^{(1-k)}\nu_{k-1}(d\boldsymbol{s})d\boldsymbol{\theta} \\
&= c_1^{-1}(\kappa)\int_{\mathcal{S}_{k-1}\times[0,2\pi)^k} s_1\exp(\kappa s_1) 2^{(1-k)}\nu_{k-1}(d\boldsymbol{s})d\boldsymbol{\theta} \\
&= c_1^{-1}(\kappa)\int_{\mathcal{S}_{k-1}\times[0,2\pi)^k} \frac{\partial}{\partial\kappa}[\exp(\kappa s_1)] 2^{(1-k)}\nu_{k-1}(d\boldsymbol{s})d\boldsymbol{\theta} \\
&= c_1^{-1}(\kappa)\frac{\partial}{\partial\kappa}\left\{2^{(1-k)}\int_{\mathcal{S}_{k-1}}\int_{[0,2\pi)^k}[\exp(\kappa s_1)]\nu_{k-1}(d\boldsymbol{s})d\boldsymbol{\theta}\right\} \\
&= c_1^{-1}(\kappa)\frac{\partial}{\partial\kappa}[c_1(\kappa)].
\end{aligned}$$

According to Abramowitz and Stegun (1964, p. 507), the *nth* derivative of confluent hypergeometric function is given by

$$\frac{d^n}{dx^n}{}_1F_1(a; b; x) = \frac{(a)_n}{(b)_n}{}_1F_1(a; b; x),$$

where $(a)_n = a(a+1)(a+2)\ldots(a+n-1), (a)_0 = 1$. Then

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{z}^*\boldsymbol{\mu}|^2) = c_1^{-1}(\kappa)\frac{\partial}{\partial\kappa}[c_1(\kappa)] &= \left[\frac{(k-1)!}{2\pi^k}\cdot\frac{1}{{}_1F_1(1; k; \kappa)}\right]\cdot\left[\frac{2\pi^k}{(k-1)!}\frac{1}{k}{}_1F_1(2; k+1; \kappa)\right] \\
&= \frac{1}{k}\frac{{}_1F_1(2; k+1; \kappa)}{{}_1F_1(1; k; \kappa)}.
\end{aligned}$$

To generalize the result to $\boldsymbol{\mu} \neq [1, 0, \ldots, 0]^\top$ arbitrary, consider a rotation matrix $\boldsymbol{R}$ unitary whose first column is $\boldsymbol{\mu}$. A suitable matrix $\boldsymbol{R}$ can be obtained using the QR

decomposition Seber (2008, p. 340), Golub and Loan (2012, p. 223-236) and Kurz and Hanebeck (2015), Kurz, Pfaff and Hanebeck (2016) . Thus, we have $\boldsymbol{\mu} = \boldsymbol{R}[1, 0, \ldots, 0]^\top$ and $\boldsymbol{R}^*\boldsymbol{\mu} = [1, 0, \ldots, 0]^\top$. In the integration procedure we use the substitution $\boldsymbol{z} = \boldsymbol{Rt}$ with $\boldsymbol{t} \in \mathbb{C}\boldsymbol{S}^{k-1}$. Then we reduce the problem to the case of $\boldsymbol{\mu} = [1, 0, \ldots, 0]^\top$. Thus

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{z}^*\boldsymbol{\mu}|^2) &= \int_{\mathbb{C}S^{k-1}} |\boldsymbol{z}^*\boldsymbol{\mu}|^2 c_1^{-1}(\kappa) \exp(\kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2) w_k(d\boldsymbol{z}) \\
&= c_1^{-1}(\kappa) \int_{\mathbb{C}\boldsymbol{S}^{k-1}} |(\boldsymbol{Rt})^*\boldsymbol{\mu}|^2 \exp(\kappa|(\boldsymbol{Rt})^*\boldsymbol{\mu}|^2) w_k(d\boldsymbol{t}) \\
&= c_1^{-1}(\kappa) \int_{\mathbb{C}\boldsymbol{S}^{k-1}} |t^*\boldsymbol{R}^*\boldsymbol{\mu}|^2 \exp(\kappa|\boldsymbol{t}^*\boldsymbol{R}^*\boldsymbol{\mu}|^2) w_k(d\boldsymbol{t}) \\
&= c_1^{-1}(\kappa) \int_{\mathbb{C}\boldsymbol{S}^{k-1}} |\boldsymbol{t}^*\boldsymbol{R}^*\boldsymbol{\mu}|^2 \exp(\kappa|\boldsymbol{t}^*\boldsymbol{R}^*\boldsymbol{\mu}|^2) w_k(d\boldsymbol{t}) \\
&= c_1^{-1}(\kappa) \int_{\mathbb{C}\boldsymbol{S}^{k-1}} |\boldsymbol{t}^*[1, 0, \ldots, 0]^\top|^2 \exp(\kappa|\boldsymbol{t}^*[1, 0, \ldots, 0]^\top|^2) w_k(d\boldsymbol{t}) \\
&= c_1^{-1}(\kappa) \int_{\mathbb{C}\boldsymbol{S}^{k-1}} s_1 \exp(\kappa s_1) 2^{(1-k)} \nu_{k-1}(d\boldsymbol{s}) d(\boldsymbol{\theta}) \\
&= c_1^{-1}(\kappa) \frac{\partial}{\partial \kappa}[c_1(\kappa)] \\
&= \frac{1}{k} \frac{{}_1F_1(2; k+1; \kappa)}{{}_1F_1(1, k, \kappa)}.
\end{aligned} \tag{C.2}$$

Therefore replaced (C.2) in (C.1) we have

$$H_S(\boldsymbol{\mu}, \kappa) = \log[c_1(\kappa)] - \frac{\kappa}{k} \frac{{}_1F_1(2, k+1, \kappa)}{{}_1F_1(1, k, \kappa)}.$$

(b) The Rényi entropy is given by

$$\begin{aligned}
H_R^\beta(\boldsymbol{\mu}, \kappa) &= \frac{1}{(1-\beta)} \log \mathbb{E}_{\boldsymbol{\theta}}\{[c_1^{-1}(\kappa) \exp(\kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2)]^{(\beta-1)}\} \\
&= \frac{1}{(1-\beta)}\{\, (1-\beta) \log c_1(\kappa) \,+\, \log\, \mathbb{E}_{\boldsymbol{\theta}}[\exp(\kappa(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}|^2)]\}.
\end{aligned} \tag{C.3}$$

We need to find the closed-form expression for expectation to the right-side in (C.3). Thus

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[\, \exp(\kappa\,(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}|^2)\,] &= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} \exp[\,\kappa(\beta-1)|\boldsymbol{z}^*\boldsymbol{\mu}|^2] c_1^{-1}(\kappa) \exp(\,\kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2\,) w_k(d\boldsymbol{z}) \\
&= \int_{\mathbb{C}\boldsymbol{S}^{k-1}} c_1^{-1}(\kappa) \exp[\,\kappa(\beta-1)\,|\boldsymbol{z}^*\boldsymbol{\mu}|^2 \,+\, \kappa|\boldsymbol{z}^*\boldsymbol{\mu}|^2\,]\, w_k(d\boldsymbol{z}) \\
&= \frac{c_1(\kappa\beta)}{c_1(\kappa)} \int_{\mathbb{C}\boldsymbol{S}^{k-1}} c_1^{-1}(\kappa\beta) \exp[\,\kappa\beta\,|\boldsymbol{z}^*\boldsymbol{\mu}|^2]\, w_k(d\boldsymbol{z}) \\
&= \frac{c_1(\kappa\beta)}{c_1(\kappa)}.
\end{aligned}$$

Then,

$$\begin{aligned} H_R^{\beta}(\boldsymbol{\mu},\kappa) &= \frac{1}{(1-\beta)}\left\{ (1-\beta)\log c_1(\kappa) + \log\left[\frac{c_1(\kappa\beta)}{c_1(\kappa)}\right]\right\} \\ &= \frac{1}{(1-\beta)}\left\{ (1-\beta)\log[c_1(\kappa)] + \log[c_1(\kappa\beta)\,] - \log[c_1(\kappa)]\right\} \\ &= \frac{1}{(1-\beta)}\left\{ \log[c_1(\kappa\beta) - \beta\log[c_1(\kappa)]\,]\right\}. \end{aligned}$$

Concluding the proof of the Theorem. ■

Proof of the Corollary 5.2.

Considering the parameter $\boldsymbol{\mu}$ fixed and $\beta \in \mathbb{R}_+ - \{1\}$ arbitrary. Making κ tend to zero, we obtain

$$\begin{aligned} \lim_{\kappa\to 0} H_R^{\beta}(\boldsymbol{\mu},\kappa) &= \lim_{\kappa\to 0}\frac{1}{(1-\beta)}\left\{\log[c_1(\beta\kappa)] - \beta\log[c_1(\kappa)]\right\} \\ &= \frac{1}{(1-\beta)}\lim_{\kappa\to 0}\left\{\log\left[\frac{2\pi^k}{(k-1)!}{}_1F_1(1;k;\beta\kappa)\right] - \beta\log\left[\frac{2\pi^k}{(k-1)!}\right]\right\} \\ &= \frac{1}{(1-\beta)}\left\{\log\left[\frac{2\pi^k}{(k-1)!}\right] - \beta\log\left[\frac{2\pi^k}{(k-1)!}\right]\right\} \\ &= \log\left[\frac{2\pi^k}{(k-1)!}\right] \end{aligned}$$

and Shannon entropy

$$\lim_{\kappa\to 0} H_S(\boldsymbol{\mu},\kappa) = \lim_{\kappa\to 0}\left\{\log c_1(\kappa) - \frac{\kappa}{k}\cdot\frac{{}_1F_1(2,k+1,\kappa)}{{}_1F_1(1,k,\kappa)}\right\} = \log\frac{2\pi^k}{(k-1)!}.$$

■

Proof of the Proposition 5.1.

For the $\mathbb{C}B$ model the entries of matrices $\boldsymbol{T}_R$ and $\boldsymbol{T}_S$, respectively, are given by

$$\begin{aligned} t_{R_i} = \frac{\partial H_R(\boldsymbol{A})}{\partial\lambda_i} &= \frac{\partial}{\partial\lambda_i}\left[\frac{1}{(1-\beta)}\left(\log\, c(\beta\boldsymbol{A}) - \beta\log c(\boldsymbol{A})\right)\right] \\ &= \frac{1}{(1-\beta)}\left[\frac{\partial\log c(\beta\boldsymbol{A})}{\partial\lambda_i} - \beta\frac{\log c(\boldsymbol{A})}{\partial\lambda_i}\right] \\ &= \frac{1}{(1-\beta)}\left[\frac{1}{c(\beta\boldsymbol{A})}\frac{\partial c(\beta\boldsymbol{A})}{\partial\lambda_i} - \frac{\beta}{c(\boldsymbol{A})}\frac{\partial c(\boldsymbol{A})}{\partial\lambda_i}\right] \\ &= \frac{1}{(1-\beta)}\left[\frac{1}{c(\beta\boldsymbol{\Lambda})}\frac{\partial c(\beta\boldsymbol{\Lambda})}{\partial\lambda_i} - \frac{\beta}{c(\boldsymbol{\Lambda})}\frac{\partial c(\boldsymbol{\Lambda})}{\partial\lambda_i}\right] \end{aligned}$$

and

$$
\begin{aligned}
t_{S_i} &= \frac{\partial H_S(\mathbf{\Lambda})}{\partial \lambda_i} \\
&= \frac{1}{c(\mathbf{\Lambda})}\frac{\partial c(\mathbf{\Lambda})}{\partial \lambda_i} - \left[\frac{1}{c(\mathbf{\Lambda})}\frac{\partial c(\mathbf{\Lambda})}{\partial \lambda_i} + \lambda_i\left(-\frac{1}{c(\mathbf{\Lambda})^2}\left(\frac{\partial c(\mathbf{\Lambda})}{\partial \lambda_i}^2\right) + \frac{1}{c(\mathbf{\Lambda})}\frac{\partial^2 c(\mathbf{\Lambda})}{\partial {\lambda_i}^2}\right)\right] \\
&= \lambda_i\left[\frac{1}{c(\mathbf{\Lambda})^2}\left(\frac{\partial c(\mathbf{\Lambda})}{\partial \lambda_i}^2\right) - \frac{1}{c(\mathbf{\Lambda})}\frac{\partial^2 c(\mathbf{\Lambda})}{\partial {\lambda_i}^2}\right].
\end{aligned}
$$

■