Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística
Programa de Pós-Graduação em Estatística

CRISTIANY DE MOURA BARROS

NEW METHODOLOGIES FOR THE REAL WATSON DISTRIBUTION

Recife
2020

CRISTIANY DE MOURA BARROS

NEW METHODOLOGIES FOR THE REAL WATSON DISTRIBUTION

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

Concentration area: Applied Statistics.

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

Recife
2020

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

B277n Barros, Cristiany de Moura
New methodologies for the real Watson distribution / Cristiany de Moura Barros. – 2020.
64 f.

Orientador: Getúlio José Amorim do Amaral.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2020.
Inclui referências e apêndices.

1. Estatística aplicada. 2. Correção de viés. I. Amaral, Getúlio José do Amorim (orientador). II. Título.

310 CDD (23. ed.) UFPE- CCEN 2020 - 61

CRISTIANY DE MOURA BARROS

NEW METHODOLOGIES FOR THE REAL WATSON DISTRIBUTION

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: 11 de fevereiro de 2020.

BANCA EXAMINADORA

Prof.(º) Getulio Jose Amorim do Amaral
UFPE

Prof.(º) Alex Dias Ramos
UFPE

Prof.(ª) Fernanda De Bastiani
UFPE

Prof.(º) Eufrásio de Andrade Lima Neto
UFPB

Prof.(º) Rodrigo Bernardo da Silva
UFPB

I dedicate this thesis first to God, to my parents,
Francisco de Holanda and Maria de Moura,
for unconditional love and Allan, for all his love.

ACKNOWLEDGMENT

I would like to thank God first for giving me the strength needed to be able to complete this step further.
To my parents, Francisco de Holanda and Maria de Moura, for the unconditional love, dedication, patience and for teaching me the values and principles that guided the formation of my personality.
To my husband Allankardec Silva Sabino, for love, understanding and patience.
To the teachers Getúlio José Amorim do Amaral and Abraão David do Nascimento Costa, for the guidance.
To the colleagues of the Suelena doctorate, Victor, Jhonnata, Rafaela, Wanessa, Leon, Isaac, Bruna, Rodrigo, Aparecida and Carlos for their friendship, affection and shared moments of joy.
To Valéria Bittencourt, for the affection, patience and friendship with which she always had she treated me and the other doctoral students.
The Examining Board for its valuable criticism and suggestions.
To CAPES, for the financial support.

ABSTRACT

Spherical data are output in various research lines. These data may be categorized as directional (when such line is directed) and axial (otherwise). Directional data can be understood as points on a sphere; while, axial data are pairs of antipodal points (i.e., opposite points) on a sphere. The Watson (W) model is often used for describing axial data. The W distribution has two parameters: the mean axis and the concentration parameter. It is known making inference under lower concentration is a hard task. First, to outperform this gap, for the W parameters, we provide an improved maximum likelihood-based estimation procedure for the W concentration parameter. In particular, we present a closed-form expression for the second-order bias according to the Cox-Snell methodology. Further, an approximated expression for the Fisher information matrix is derived as well. To quantify the performance of the our proposal, a Monte Carlo study is made. Results indicate that our estimation procedure is suitable to obtain more accurate estimates for the W concentration parameter. Second, aims to study a natural extension of the minimum distance estimators discussed by Cao et al. (1994).More precisely, under the assumption of Watson directional distribution, to produce hypotheses and point statistical inference procedures, as well as goodness, and to propose mathematical antecedents for the statistical method based on the minimum distance of L^2 for the Watson model. Third, based on Renyi divergence, we propose two hypothesis tests to verify whether two samples come from populations with the same concentration parameter. The results of synthetic and real data indicate that the proposed tests can produce good performance on Watson data. The small sample behavior of the proposed estimators is using Monte Carlo simulations. An application is illustrated with real data sets.

Keywords: Bias correction. Watson distribution. Axial data. Minimum L^2 distance. Bootstrap and permutation tests.

RESUMO

Dados esféricos são produzidos em várias linhas de pesquisa. Esses dados podem ser categorizado como direcional (quando essa linha é direcionada) e axial (caso contrário). Dados direcionais podem ser entendidos como pontos em uma esfera; enquanto, dados axiais são pares de pontos antipodais (isto é, pontos opostos) em uma esfera. O modelo Watson (W) é frequentemente usado para descrever dados axiais. O W distribuição tem dois parâmetros: o eixo médio e a concentração parâmetro. Sabe-se que fazer inferência sob menor concentração é uma tarefa difícil. Primeiro, para superar essa lacuna, para os parâmetros W, fornecemos um procedimento melhorado de estimativa baseada em máxima verossimilhança para o W parâmetro de concentração. Em particular, apresentamos uma expressão de forma fechada para o viés de segunda ordem, de acordo com a metodologia de Cox-Snell. Além disso, uma expressão aproximada para a matriz de informações de Fisher é derivado também. Para quantificar o desempenho de nossa proposta, um estudo de Monte Carlo é feito. Os resultados indicam que nosso procedimento de estimativa é adequado para obter estimativas mais precisas para o parâmetro de concentração W. Second, realizamos o estudo sobre uma extensão natural dos estimadores de mínima distância discutidos por Cao et al. (1994). Mais precisamente, sob o pressuposto de distribuição direcional Watson, para produzir hipóteses e apontar procedimentos de inferência estatísticas, bem como bondade, e propor antecedentes matemáticos para o método estatístico com base na distância mínima de L^2 para o modelo de Watson. Terceiro, com base na divergência de Rényi, propomos dois testes de hipótese para verificar se duas amostras provêm de populações com a mesma concentração parâmetro. Os resultados de dados sintéticos e reais indicam que a proposta de testes podem produzir um bom desempenho nos dados da Watson. O comportamento em pequenas amostras dos estimadores propostos está usando simulações de Monte Carlo. Uma aplicação é ilustrada com conjuntos de dados reais.

Palavras-chave: Correção de viés. Distribuição Watson. Dados axiais. Distância mínima L^2. Bootstrap e teste de permutação.

CONTENTS

1 INTRODUCTION

Statistical analysis in the unitary sphere is not easy task, the beauty of the probabilistic models make it seem simpler than it is pratice. This difficulty usually results from the complicated normalization of the constants associated with directional distributions. However, due to their powerful modeling ability, hypersphere distributions continue to encounter numerous applications, see for example, Mardia and Jupp (2000).

The most well-known directional distribution is the Von-Mises-Fisher distribution (VMF), which models data concentrated around a mean (direction). But for the data that has an additional structure it is necessary to define what additional structure is, and this distribution may not be adequate: in particular for axially symmetric data it is more convenient to approach the Watson distribution, Watson (1965), which is the focus this thesis. Three main reasons motivate our study of the multivariate distribution. First it is fundamental for the direction statistics, second it has not received much attention for the analysis of modern data involving large data, and it is a procedure for analysis of the genetic expression Dhillon (2003).

One reason may be that the traditional domains of directional statistics are three-dimensional and two-dimensional axes, for example, circles or spheres. The Watson distribution is formed by two parameters: μ and κ, where κ is known as normalization constant and its density will be defined in (2.1).

A basic set of summary statistics in exploratory data analysis consists of the median of the sample and the extremes (also known as "hinges"), which are the approximate sample quantiles. In a two-dimensional plane, the geodesic is the shortest distance that joins two points such that, for small variations in the shape of the curve. The representation of the geodesic in a plane represents the projection of a maximum circle on a sphere. Thus, either on the surface of a sphere or deformed in a plane, the line is a curve, since the shortest possible distance between two points can only be curved, since a line would necessarily need to remain always in a plane, to be the shortest distance between points. From the practical point of view, in most cases, the geodesic is the shortest curve that joins two points.

In a "flat geometry" (Euclidean space), this curve is a straight segment, but in "curved geometries" (riemaniana geometry), much used for example in General Relativity Theory, the shortest distance curve between two points it may not be a straight line. To understand this, let us take as an example the curvature of the globe and its continents. If we draw a line connecting two capitals of different continents, we will notice that the line is not straight, but an arc of the maximum circle, however, if the distance between the two cities is small, the line covering the segment of the maximum circle will be really a straight. In general relativity, geodesics describe the motion of point particles under the influence of gravity. In particular, the path taken by a falling rock, a satellite orbit, or in the form of a planetary orbit are all geodetic in curve-time-space.

Thesis Objective

The aim of the thesis is to propose a closed-form expression for the second-order bias according to the Cox-Snell methodology for the real Watson. Then perform study a natural extension of the minimum distance estimators discussed by Cao et al.(1994). More precisely, under the assumption of Watson directional distribution, to produce hypotheses and point statistical inference procedures, as well as goodness, and to propose mathematical antecedents for the statistical method based on the minimum distance of L^2 for this distribution. Finally, based on Rényi divergence, propose two hypothesis tests to verify whether two samples come from populations with the same concentration parameter.

Thesis Organization

In addition to the introductory chapter, this thesis consists of four more chapters. In Chapter 2 we focused on data analysis on the sphere and Watson distribution. We present a general review on the data in the sphere, on the Watson distribution discussing its main characteristics and estimation of its parameters, then propose a closed-form expression for the second-order bias according to the Cox-Snell methodology for the real Watson.

In Chapter 3, aims to study a natural extension of the minimum distance estimators discussed by Cao et al. (1994), to produce hypotheses and point statistical inference procedures, as well as goodness, and to propose mathematical properties for the statistical method based on the minimum distance of L^2 for this distribution. In this same chapter we present new properties about the Watson distribution.

In Chapter 4, we propose two-samples divergence-based hypothesis tests involving concentration parameter of the Watson distribution; i.e., two statistical procedures to identify if two axial samples are similarly concentrated used permutation and bootstrap tests for two sample problems of axial data analysis. Finally in Chapter 5, we summarize the main contributions of this thesis.

The following figures 1, 2 and 3 are examples of axial data sets. They quantified simple are represented by latitude θ and longitude ϕ on three populations, based on these angles presented vectors on the unit sphere can be obtained by the transformation: $x = \sin(\theta)\cos(\phi)$, $y = \sin(\theta)\sin(\phi)$ and $z = \cos(\theta)$. Figure 1 presents the graph of pole positions from the study of paleomagnetic soils of New Caledonia by Fisher, Lewis and Embleton (1987).

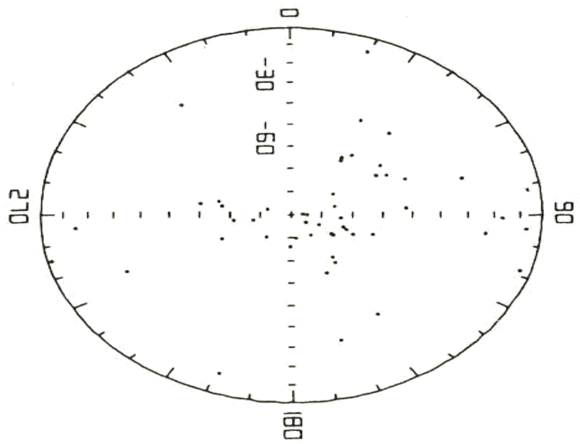


Figure 1 – Positions of poles from the study of paleomagnetic soils of New Caledonia.

The second and third samples, both refer to orientations of axial plane cleavage surfaces of turbidite ordovician folds, composed by the variables: diving and diving direction, denoted latitude θ and longitude ϕ , respectively. Figure 2 shows the projection plot of 72 poles for axial plane cleavage surfaces. Figure 3 shows the projection plot of 75 for axial plane cleavage surfaces.

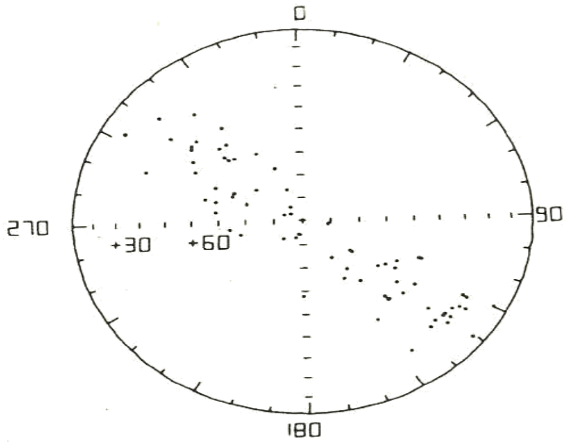


Figure 2 – Projection of 72 poles for axial plane cleavage surfaces.

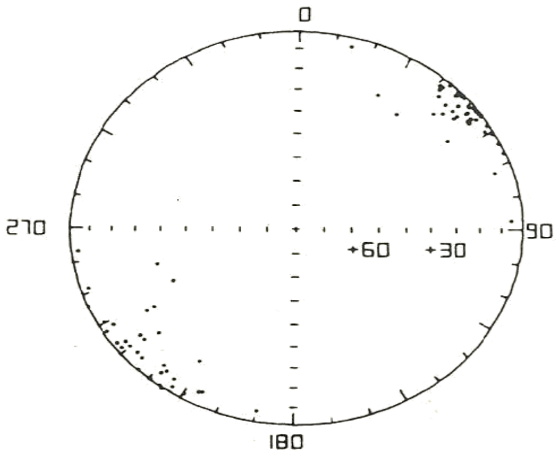


Figure 3 – Projection of 75 poles for axial plane cleavage surfaces.

Computational support

As for the computational part, was used R software, which is a language and an environment for computing statistics and for the preparation of high quality grades. R offers a wide range of statistical techniques and graphics. The fact that it is a free software is allowed contributions of new features through the creation of packages. Documentation and tutorials are available at http://www.r-project.org.

2 BIAS-CORRECTED ESTIMATION FOR THE REAL WATSON

2.1 Introduction

Spherical data are commonly represented by a straight line within normed space Fisher, Embleton and Lewis (1993). They are output in various research lines: theory of shape Small (1996), equatorial distributions on a sphere Watson (1965) and discordancy tests for samples on the sphere Best and Fisher (1986). These data may be categorized as directional (when such line is directed) or axial (otherwise). Directional data can be understood as points on a sphere; while, axial data are pairs of antipodal points (i.e., opposite points) on a sphere. The Watson (W) model is often used for describing axial data Fisher, Embleton and Lewis (1993).

The W model can be seen as an extension of the well-known Von-Mises-Fisher distribution Fisher, Embleton and Lewis (1993). While the last supposed that axial data are concentrated around a mean (direction), the W distribution may also describe axes dispersed over mean and its variability depend a concentration parameter. The W distribution has two parameters: the mean axis and the concentration parameter (say κ). The W flexibility lies in κ regard to both degree and kind of axial data dispersion. The concentration is directly proportional to the values of $|\kappa|$. If κ is positive, the distribution is bipolar. If κ is negative, the distribution is girdle.

The maximum likelihood (ML) estimation has been widely used mainly due to its good asymptotic properties, such as consistency and convergence in distribution to the Gaussian law. In contrast, the ML estimator is often biased with respect to the true parameter value. The former has bias of order $O(N^{-1})$, where N is the sample size and $O(\cdot)$ is the Landau notation to represent order. The fact of the bias value be negligible comparatively to the standard error (which has order $O(N^{-1/2})$) becomes the previous phenomenon unimportant. However, such biases can be expressive before small or moderate sample sizes. As a solution, analytic expressions for the ML-estimator bias are required to derive a more accurate corrected estimator for finite sample sizes see Cordeiro and Cribari (2014).

Several bias-corrected methods for models in directional data have been proposed in the literature. Some of them are improved ML estimators for the von Mises-Fisher concentration parameters Best and Fisher (1981) and the parameters of the complex Bingham distribution Dore *et. al.* (2016).

In this paper, we propose an improved estimator for κ in order to correct the ML bias, mainly under lower concentration axial data, which impose often a more expressive bias.

To that end, we follow the methodology proposed by Cox and Snell (1968) in terms of the second-order bias expression of (ML) estimator. Closed-form approximated expressions for the Fisher Information Matrix (FIM) and associated bias are derived for the parameters and a new estimator for κ is proposed. Subsequently, a simulation study is made to quantify the performance of our proposal. Results evidence in favor of the our proposal like an efficient tool to measure concentration over axial data.

This chapter is organized as follows. In Section 2.2, the Watson distribution is introduced. Section 2.3 presents an outline on the Cox-Snell correction and the proposal of an estimator for κ. In Section 2.4, a performance study is carried out. Finally, Section 2.5 summarizes the main conclusions of this chapter.

2.2 The Watson Distribution

The Watson distribution has support on the unit sphere, say $\mathbb{S}^{p-1} = \{\boldsymbol{x} \in \mathbb{R}^p \colon \boldsymbol{x}^\top \boldsymbol{x} = 1\}$, and probability density function (pdf) given by

$$\begin{aligned} f(\boldsymbol{x};\boldsymbol{\mu},\kappa) &= \frac{\Gamma(p/2)}{2\,\pi^{p/2}} M\left(\frac{1}{2},\frac{p}{2},\kappa\right)^{-1} \exp\{\kappa\,(\boldsymbol{\mu}^\top \boldsymbol{x})^2\} \\ &= \frac{\Gamma(p/2)}{2\,\pi^{p/2}} M\left(\frac{1}{2},\frac{p}{2},\kappa\right)^{-1} \exp\{\kappa\,\cos(\theta(\boldsymbol{x},\boldsymbol{\mu}))^2\}, \end{aligned} \tag{2.1}$$

for $\pm\boldsymbol{x},\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ and $\kappa \in \mathbb{R}$, where $\theta(\boldsymbol{x},\boldsymbol{\mu}) = \arccos[\boldsymbol{\mu}^\top \boldsymbol{x}]$ represents the angle between the dominant axis and a possible outcome belonging to $\mathbb{S}^{p-1}$ and

$$M(a,b,z) \;=\; \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 \mathrm{e}^{z\,u}\, u^{a-1}\,(1-u)^{b-a-1}\,\mathrm{d}u \;=\; \sum_{n=0}^{\infty} \frac{a^{(n)}\, z^n}{b^{(n)}\, n!}$$

is the Kummer function Abramowitz and Stegun (1994), where $a^{(0)} = 1$ and $a^{(n)} = a\,(a+1)\,(a+2)\,\cdots\,(a+n-1)$. This case is denoted as $\boldsymbol{x} \sim \boldsymbol{W_p}(\boldsymbol{\mu},\kappa)$, where $\pm\boldsymbol{\mu}$ indicates the dominant axis and κ is the concentration parameter. Moreover, the Watson model is rotationally symmetric about $\pm\boldsymbol{\mu}$. If $\kappa < 0$, this distribution has a mode around the equator at 90º to the axis $\pm\boldsymbol{\mu}$ (this situation is classified as *girdle form*). On the other hand, for $\kappa > 0$, the distribution is bipolar with modes at $\pm\boldsymbol{\mu}$ (denoted as *bipolar form*). When $\kappa = 0$ the uniform distribution is obtained. Further, this distribution also satisfies the following properties:

- The Watson density is rotationally symmetric around $\pm\boldsymbol{\mu}$;
- If $\boldsymbol{x} \sim \boldsymbol{W_p}(\boldsymbol{\mu},\kappa)$ and $\boldsymbol{A}$ an orthogonal matrix, then $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \sim \boldsymbol{W_p}(\boldsymbol{A}\boldsymbol{\mu},\kappa)$ Mardia and Jupp (2000),
- The following high concentration approximations Mardia and Jupp (2000) hold for the W model:

$$2\kappa\{1-(\boldsymbol{\mu}^\top \boldsymbol{x})^2\} \sim \chi^2_{p-1}, \;\; \kappa \to \infty,$$

in the bipolar case, and similarly

$$2|\kappa|\{1-(\boldsymbol{\mu}^\top \boldsymbol{x})^2\} \sim \chi^2_{1}, \;\; \kappa \to -\infty,$$

in girdle case.

2.2.1 Maximum likelihood estimation

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \in \mathbb{S}^{p-1}$ be a random sample from $\boldsymbol{x} \sim \boldsymbol{W_p}(\boldsymbol{\mu}, \kappa)$. The log-likelihood function at $\boldsymbol{\theta}^\top = (\boldsymbol{\mu}^\top, \kappa)$ is given by

$$\ell(\boldsymbol{\theta}; \boldsymbol{X}) \;=\; n\left[\kappa\, \boldsymbol{\mu}^\top\, \mathbf{S}\, \boldsymbol{\mu} - \log\, M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right) + \gamma\right], \tag{2.2}$$

where $\boldsymbol{X} = [\boldsymbol{x_i}, \ldots, \boldsymbol{x_n}]^\top$ is the observed sample and $\boldsymbol{S} = n^{-1}\sum_{i=1}^n \boldsymbol{x_i}\boldsymbol{x_i}^\top$ is *the scattering matrix of the sample* or *of sampling orientation matrix* and $\gamma = \log\left[\frac{\Gamma(p/2)}{2\pi^{p/2}}\right]$ is a constant term, often suppressed in the inferential process.

The maximum likelihood estimate (MLE) for $\boldsymbol{\theta} = (\kappa, \boldsymbol{\mu}^\top)$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}; \dot{\boldsymbol{X}}),$$

where $\boldsymbol{\Theta}$ represents the associated parametric space.

Specifically, Sra and Karp (2013) showed the MLE for $\boldsymbol{\mu}$, $\widehat{\boldsymbol{\mu}}$, is given by

$$\widehat{\boldsymbol{\mu}} = \begin{cases} \boldsymbol{s}_1, & \text{if } \hat{\kappa} > 0 \text{ (bipolar form)}, \\ \boldsymbol{s}_p, & \text{if } \hat{\kappa} < 0 \text{ (girdle form)}, \end{cases}$$

where $\boldsymbol{s}_1$ and $\boldsymbol{s}_p$ represent the normalized eigenvectors associated with the eigenvalues of $\boldsymbol{S}$. On the other hand, the MLE for the concentration parameter κ is obtained by solving the following non-linear equation:

$$g\left(\frac{1}{2}, \frac{p}{2}; \hat{k}\right) \equiv \frac{M'\left(\frac{1}{2}, \frac{p}{2}, \hat{\kappa}\right)}{M\left(\frac{1}{2}, \frac{p}{2}, \hat{\kappa}\right)} = \hat{\boldsymbol{\mu}}^\top\, \boldsymbol{S}\, \hat{\boldsymbol{\mu}} \equiv r \qquad (0 \le r \le 1),$$

where $M'(\cdot,\cdot,\cdot)$ is the derivative of $M(\cdot,\cdot,\cdot)$ with respect to κ given by

$$M'(a, b, \kappa) = \frac{a}{b} M(a+1, b+1, \kappa).$$

Thus, two conditions need to be satisfied to obtain the MLE for κ: (i) $\lambda_1 > \lambda_2$ for $\kappa > 0$ and (ii) $\lambda_{p-1} > \lambda_p$ for $\kappa < 0$, where $\lambda_1, \cdots, \lambda_p$ are eigenvalues of $\boldsymbol{S}$. As discussed by Sra and Karp (2013), the Equation can rewritten as

$$g\left(\frac{1}{2}, \frac{p}{2}; \hat{\kappa}\right) \;=\; \lambda_1 \quad \text{or} \quad g\left(\frac{1}{2}, \frac{p}{2}; \hat{\kappa}\right) \;=\; \lambda_p.$$

From now on, we assume that MLEs for κ are defined as solutions of

$$g(a,c;k) \equiv \frac{M'(a,c,k)}{M(a,c,k)} = r, \qquad c > a > 0, \quad 0 \leq r \leq 1,$$

where r is the smallest or largest eigenvalue from the sampling orientation matrix. The MLE for $\boldsymbol{\mu}$ is the eigenvector corresponding to the rth eigenvalue.

Even though $\widehat{\boldsymbol{\mu}}$ is easily obtained, it is hard to compute. So, Sra and Karp (2013) have derived asymptotic approximations for $\hat{\kappa}$ given by
$\kappa(r) = \frac{-a}{c} + (c-a-1) + \frac{(c-a-1)(1+a)}{a} r + O(r^2), \quad r \to 0,$
$\kappa(r) = \left(r - \frac{a}{c}\right)\left\{\frac{c^2(1+c)}{a(c-a)} + \frac{c^3(1+c)^2(2a-c)}{a^2(c-a)^2(c+2)}\left(r - \frac{a}{c}\right) + O\left(\left(r - \frac{a}{c}\right)\right)\right\}, r \to \frac{a}{c},$ and
$\kappa(r) = \frac{c-a}{1-r} + 1 - a + \frac{(a-1)(a-c-1)}{c-a}(1-r) + O((1-r)^2), \quad r \to 1.$ In this case, the observed information matrix, say $\boldsymbol{J}(\boldsymbol{\theta})$, is defined as

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2 \ell}{\partial^2 \mu_1} & \cdots & \frac{\partial^2 \ell}{\partial \mu_1 \partial \mu_p} & \frac{\partial^2 \ell}{\partial \mu_1 \partial \kappa} \\ \frac{\partial^2 \ell}{\partial \mu_2 \partial \mu_1} & \cdots & \frac{\partial^2 \ell}{\partial \mu_2 \partial \mu_p} & \frac{\partial^2 \ell}{\partial \mu_2 \partial \kappa} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial \kappa \partial \mu_1} & \cdots & \frac{\partial^2 \ell}{\partial \kappa \partial \mu_p} & \frac{\partial^2 \ell}{\partial^2 k} \end{pmatrix}. \tag{2.3}$$

Based on (2.3), under certain regularity conditions, one can define the FIM, say $\boldsymbol{K}(\boldsymbol{\theta})$, as $\boldsymbol{K}(\boldsymbol{\theta}) = \mathrm{E}[-\boldsymbol{J}(\boldsymbol{\theta})]$. To define $\boldsymbol{K}(\boldsymbol{\theta})$, it is necessary two results: (i) obtaining the 2nd derivative of ℓ with respect to $\boldsymbol{\mu}$ and κ (Barros *et. al.* 2016) and (ii) determining the components of $\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top)$ in $\boldsymbol{K}(\boldsymbol{\theta})$. For $\boldsymbol{x}$ following the real Bingham distribution having parameter $\boldsymbol{\Sigma}$, it holds that we use an approximation proposed by (Kume and Walker 2014):

$$\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) \approx \frac{\mathbb{I}_p}{p+1} + \frac{2\boldsymbol{\Sigma}}{(p+1)(p+3)} + \frac{2\,\mathbb{I}_p \operatorname{tr}(\boldsymbol{\Sigma})}{(p+1)^2(p+3)}, \tag{2.4}$$

where $\operatorname{tr}(\cdot)$ is the trace operator and $\mathbb{I}_{\boldsymbol{p}}$ is the identity matrix at the order p. Since the W model may be rewritten from the Bingham making $\boldsymbol{\Sigma} = (\mathbb{I}_{\boldsymbol{p}} - 2\kappa\boldsymbol{\mu}\boldsymbol{\mu}^\top)^{-1}$, Mardia and Jupp (2000), we use this result to obtain (ii), see appendix *A*.

2.3 Corrected estimation for the concentration

The methodology proposed by Cox and Snell (1968) is shortly reviewed in this section. The notation is defined as follows:

$$U_i = \frac{\partial}{\partial \theta_i}\ell(\boldsymbol{\theta}), U_{ij} = \frac{\partial^2}{\partial \theta_i \partial \theta_j}\ell(\boldsymbol{\theta}), U_{ijk} = \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k}\ell(\boldsymbol{\theta}),$$

for $i, j, k = 1, 2, \ldots, p$. Moreover, the cumulants for the log-likelihood derivatives are given by

$$\kappa_{ij} = \mathrm{E}(U_{ij}), \qquad \kappa_{i,j} = \mathrm{E}(U_i U_j), \qquad \kappa_{ijk} = \mathrm{E}(U_{ijk}), \qquad \kappa_{i,jk} = \mathrm{E}(U_i U_{jk}).$$

Consequently, the elements of the FIM at $\boldsymbol{\theta}$ are κ_{ij}. It is known, under the regularity conditions that $\kappa_{i,j} = -\kappa_{ij}$. The entries of the inverse of FIM as κ^{ij}.

Furthermore, the first derivatives of the cumulants κ_{ij} with respect to the W parameters are denoted by:

$$\kappa_{ij}^{(k)} = \frac{\partial}{\partial \theta_k} \kappa_{\theta_i \theta_j},$$

for $i, j, k = 1, 2, \ldots, p$, in case for $p = 3$.

2.3.1 Cox-Snell Methodology

Cox and Snell (1968) derived a formula for the second-order bias of the ML estimator for $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^\top$. According to these authors, if $\hat{\theta}_a$ is the ML estimator for θ_a in $\boldsymbol{\theta}$, the following expression for the bias of $\boldsymbol{\theta}$ is:

$$\mathrm{B}(\hat{\boldsymbol{\theta}}_a) = \mathbb{E}(\hat{\boldsymbol{\theta}}_a) - \theta_a = \sum_{r,s,t} \kappa^{ar} \kappa^{st} \left(\kappa_{rs}^{(t)} - \frac{1}{2} \kappa_{rst} \right) + \sum_{r=2}^{\infty} O(n^{-r}), \tag{2.5}$$

where $r, s, t = 1, \cdots, p$. A corrected ML estimator, say $\tilde{\boldsymbol{\theta}}_a$, may be given by $\tilde{\boldsymbol{\theta}}_a = \hat{\boldsymbol{\theta}}_a - \hat{\boldsymbol{B}}(\hat{\boldsymbol{\theta}}_a)$, where $\hat{\boldsymbol{B}}(\hat{\boldsymbol{\theta}}_a)$ indicates the bias $\boldsymbol{B}(\hat{\boldsymbol{\theta}}_a)$ evaluated at $\hat{\boldsymbol{\theta}}_a$. It is know $\mathbb{E}(\hat{\boldsymbol{\theta}}_a) = \boldsymbol{\theta_a} + \boldsymbol{O}(n^{-1})$, while $\mathbb{E}(\tilde{\boldsymbol{\theta}}_a) = \boldsymbol{\theta_a} + \boldsymbol{O}(n^{-2})$. Thus, $\tilde{\boldsymbol{\theta}}_a$ has better asymptotic properties than $\hat{\boldsymbol{\theta}}_a$.

It what follows, the expression (2.5) is applied to correct the ML estimator for the W concentration parameter. It is known $\mathbb{E}(\boldsymbol{x}) = 0$ and the random number generator for W variables depends only of κ according to Kim-Hung and Carl (1993). From the last note, we focus only on the improved estimation of κ.

2.3.2 Expression for the second-order bias

As first contribution of this chapter, we derive a closed-form expression for W FIM, given in Corollary 2.1.

Corollary 2.1. *The components of the unit (for sample size n=1) FIM are given by*

$$\kappa_{11} \approx \frac{13\kappa - 28\kappa^2\mu_3^2 - 28\kappa^2\mu_2^2 - 20\kappa^2\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)}, \quad \kappa_{12} \approx \kappa_{21} \approx \frac{\kappa^2\mu_1\mu_2}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{13} \approx \kappa_{31} \approx \frac{\kappa^2\mu_1\mu_3}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)}, \quad \kappa_{14} \approx \kappa_{41} = \frac{13\mu_1 - 20\kappa\mu_1\mu_3^2 - 20\kappa\mu_1\mu_2^2 - 20\kappa\mu_1^3}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{22} \approx \frac{13\kappa - 28\kappa^2\mu_3^2 - 20\kappa^2\mu_2^2 - 28\kappa^2\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)}, \quad \kappa_{23} \approx \kappa_{32} \approx \frac{\kappa^2\mu_2\mu_3}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{24} \approx \kappa_{42} \approx \frac{13\mu_2 - 20\kappa\mu_1^2\mu_2 - 20\kappa\mu_2\mu_3^2 - 20\kappa\mu_2^3}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)}, \quad \kappa_{33} \approx \frac{13\kappa - 20\kappa^2\mu_3^2 - 28\kappa^2\mu_2^2 - 28\kappa^2\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{34} \approx \kappa_{43} \approx \frac{13\mu_3 - 20\kappa\mu_3^3 - 20\kappa\mu_1^2\mu_3 - 20\kappa\mu_2^2\mu_3}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{44} \approx \frac{0.5^2(2.5)M(1.5, 2.5, \kappa)^2 - 0.5(1.5)^2 M(0.5, 1.5, \kappa)M(2.5, 3.5, \kappa)}{1.5^2(2.5)M(0.5, 1.5, \kappa)^2}.$$

and $\kappa_{ii} \approx \frac{13\kappa - 20\kappa^2 - 28\kappa^2\left(\sum_{j\neq i}\boldsymbol{\mu}_j^2\right)}{24\left[1-2\kappa\left(\sum_{h=1}^{3}\boldsymbol{\mu}_h^2\right)\right]}$,

$\kappa_{ij} \approx \frac{\kappa^2\boldsymbol{\mu}_i\boldsymbol{\mu}_j}{3\left[1-2\kappa\left(\sum_{h=1}^{3}\boldsymbol{\mu}_h^2\right)\right]}$, i,j$= 1, 2, 3$,

$\kappa_{i4} \approx \frac{13\boldsymbol{\mu}_i - 20\kappa\boldsymbol{\mu}_i^3 - 20\kappa\boldsymbol{\mu}_i\left(\sum_{j\neq i}\boldsymbol{\mu}_j^2\right)}{3\left[1-2\kappa\left(\sum_{h=1}^{3}\boldsymbol{\mu}_h^2\right)\right]}$, i,j$= 1, 2, 3.$

An outline of the proof these results is given in appendix A.

Based in Corollary 2.1, the term κ_{ijj} of (2.5) for the W model is derived. The expected value of the third derivatives obtained from the FIM with respect to the W parameters are expressed by

$$\kappa_{11\bullet} \approx \kappa_{12\bullet} \approx \kappa_{13\bullet} \approx \kappa_{\bullet 44} \approx \kappa_{21\bullet} \approx \kappa_{22\bullet} \approx \kappa_{23\bullet} \approx \kappa_{31\bullet} \approx \kappa_{32\bullet} \approx \kappa_{33\bullet} \approx \kappa_{4\bullet 4} \approx \kappa_{44\bullet} \approx 0,$$

$$\kappa_{114} \approx \kappa_{141} \approx \kappa_{411} \approx n\frac{13 - 28\kappa\mu_3^2 - 28\kappa\mu_2^2 - 20\kappa\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{124} \approx \kappa_{142} \approx \kappa_{214} \approx \kappa_{241} \approx \kappa_{412} \approx \kappa_{421} \approx n\frac{\kappa\mu_1\mu_2}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{134} \approx \kappa_{143} \approx \kappa_{314} \approx \kappa_{341} \approx \kappa_{413} \approx \kappa_{431} \approx n\frac{\kappa\mu_1\mu_3}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{224} \approx \kappa_{242} \approx \kappa_{422} \approx n\frac{13 - 28\kappa\mu_3^2 - 20\kappa\mu_2^2 - 28\kappa\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{234} \approx \kappa_{243} \approx \kappa_{324} \approx \kappa_{342} \approx \kappa_{423} \approx \kappa_{432} \approx n\frac{\kappa\mu_2\mu_3}{3(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)},$$

$$\kappa_{334} \approx \kappa_{343} \approx \kappa_{433} \approx n\frac{13 - 20\kappa\mu_3^2 - 28\kappa\mu_2^2 - 28\kappa\mu_1^2}{24(1 - 2\kappa{\mu_3}^2 - 2\kappa{\mu_2}^2 - 2\kappa{\mu_1}^2)}$$

and

$$\kappa_{444} \approx n\frac{(1.5)M(0.5, 1.5, \kappa)M(1.5, 2.5, \kappa)M(2.5, 3.5, \kappa) - (2.5)M(2.5, 3.5, \kappa)M(1.5, 2.5, \kappa)}{1.5^2(2.5)M(0.5, 1.5, \kappa)^3}.$$

Let $\boldsymbol{x} \sim \boldsymbol{W_p}(\boldsymbol{\mu}, \kappa)$ and $\hat{\kappa}$ be the ML estimator for κ based on a random sample of n points from $\boldsymbol{x}$. Then second-order bias of $\tilde{\kappa}$ is given by

Corollary 2.2. *Consider $\boldsymbol{X}$ following the Watson distribution having the concentration $\tilde{\kappa}$. An expression for the bias is given*

$$
\begin{aligned}
Bias(\tilde{\kappa}) = \frac{1}{n}\Big\{ & \frac{4}{D_1}\left[\kappa^{41}\kappa^{11}N_1 + \kappa^{41}\kappa^{12}N_2 + \kappa^{41}\kappa^{13}N_3 + \kappa^{42}\kappa^{21}N_{19} + \kappa^{42}\kappa^{12}N_{20}\right. \\
& \left. + \kappa^{42}\kappa^{23}N_{21} + \kappa^{43}\kappa^{31}N_{31} + \kappa^{43}\kappa^{32}N_{32}\kappa^{43}\kappa^{33}N_{33}\right] + \left(\kappa^{41}\kappa^{21} + \kappa^{42}\right. \\
& \left. \times\kappa^{11}\right)\frac{8N_6}{D_1} + \left(\kappa^{41}\kappa^{22} + \kappa^{42}\kappa^{12}\right)\frac{8N_7}{D_1} + \left(\kappa^{41}\kappa^{23} + \kappa^{41}\kappa^{32} + \kappa^{42}\kappa^{13}\right. \\
& \left. +\kappa^{42}\kappa^{31} + \kappa^{43}\kappa^{12} + \kappa^{43}\kappa^{21}\right)\frac{8N_8}{D_1} + \left(\kappa^{41}\kappa^{31} + \kappa^{43}\kappa^{11}\right)\frac{8N_{11}}{D_1} + \left(\kappa^{41}\right. \\
& \left. \times\kappa^{33} + \kappa^{43}\kappa^{13}\right)\frac{8N_{12}}{D_1} + \frac{2\left[\kappa^{41}\kappa^{14}+(\kappa^{41}\kappa^{41}+\kappa^{44}\kappa^{11})N_{15}\right]}{2D_1} \\
& + \frac{2\left[\kappa^{42}\kappa^{24}N_{22}+(\kappa^{42}\kappa^{42}+\kappa^{44}\kappa^{22})N_{28}+\kappa^{43}\kappa^{34}N_{34}+\kappa^{43}\kappa^{43}N_{37}\right]}{2D_1} \\
& + \frac{\kappa^{44}\kappa^{33}N_{37}}{2D_1D_2} - \frac{\left[\kappa^{41}\kappa^{14}+(\kappa^{41}\kappa^{41}+\kappa^{44}\kappa^{11})N_5+\left(\kappa^{42}\kappa^{24}+\kappa^{42}\kappa^{22}\right)\right]}{2D_1D_2} \\
& - \frac{\left[\kappa^{44}\kappa^{22}N_{23}+(\kappa^{43}\kappa^{34}+\kappa^{43}\kappa^{43}+\kappa^{44}\kappa^{33})N_{35}\right]}{2D_1D_2} + \frac{8\kappa^{41}\kappa^{24}N_9}{D_1} \\
& + \frac{\left\{8\left[\kappa^{42}\kappa^{14}N_9+(\kappa^{41}\kappa^{34}+\kappa^{43}\kappa^{14})N_{13}+(\kappa^{42}\kappa^{34}+\kappa^{43}\kappa^{24})N_{26}\right]\right\}}{D_1} \\
& + \frac{\left\{12\left[(\kappa^{44}\kappa^{12}+\kappa^{44}\kappa^{21})N_{16}+(\kappa^{41}\kappa^{43}+\kappa^{44}\kappa^{13}+\kappa^{44}\kappa^{31})N_{17}\right]\right\}}{D_1} \\
& + \frac{\left\{12\left[(\kappa^{42}\kappa^{43}+\kappa^{43}\kappa^{42+\kappa^{44}\kappa^{23}+\kappa^{44}\kappa^{32}})N_{29}+\kappa^{43}\kappa^{41}N_{36}\right]\right\}}{D_1} \\
& - \frac{4\left[(\kappa^{41}\kappa^{24}+\kappa^{42}\kappa^{14}+\kappa^{41}\kappa^{42}+\kappa^{42}\kappa^{41}+\kappa^{44}\kappa^{22}+\kappa^{44}\kappa^{21})N_{10}\right]}{D_2} \\
& - \frac{4\left[(\kappa^{41}\kappa^{34}+\kappa^{43}\kappa^{14}+\kappa^{41}\kappa^{43}+\kappa^{44}\kappa^{13}+\kappa^{44}\kappa^{31}+\kappa^{43}\kappa^{41})N_{14}\right]}{D_2} \\
& - \frac{4\left[(\kappa^{42}\kappa^{34}+\kappa^{43}\kappa^{24}+\kappa^{42}\kappa^{43}+\kappa^{42}\kappa^{42}+\kappa^{44}\kappa^{23}+\kappa^{44}\kappa^{32})N_{27}\right]}{D_2} \\
& +\left(\kappa^{42}\kappa^{32} + \kappa^{43}\kappa^{22}\right)\frac{8N_{24}}{D_1} + \left(\kappa^{42}\kappa^{33} + \kappa^{43}\kappa^{23}\right)\frac{8N_{25}}{D_1} + \kappa^{42}\kappa^{44}\frac{6N_{30}}{D_1} \\
& +\kappa^{43}\kappa^{44}\left(\frac{6N_{38}}{D_1}\right) + \kappa^{44}\kappa^{14}\left(\frac{6N_{39}}{D_1}\right) + \kappa^{44}\kappa^{24}\left(\frac{6N_{40}}{D_1}\right) + \kappa^{44}\kappa^{34} \\
& \times\left(\frac{6N_{41}}{D_1}\right) + \kappa^{44}\kappa^{44}\left(\frac{2N_{42}-N_{43}}{2D_3}\right)\Big\},
\end{aligned}
$$

where $\mathbf{N}_i$, $\mathbf{N}_{ij}$ and $\mathbf{D}_i$ are given in appendix B.

2.4 Numeric Results

The purpose of this section is to quantify the performance of MLE for κ comparatively to its corrected version. A simulation experiment is performed as follows. The number of Monte Carlo samples is 5000. For each Monte Carlo sample are computed the bias and the mean square error (MSE). We considered both bipolar ($\kappa = 1, 3, 5, 6$) and girdle ($\kappa = -6, -5, -3, -1$) cases and sample sizes $n = 20, 30, 50$. These cases are illustrated in Figure 4.

Table 1 shows values for bias and MSE at the bipolar case. The bias and the MSE are reduced when the sample size increases. It is expected according to the MLE asymptotic

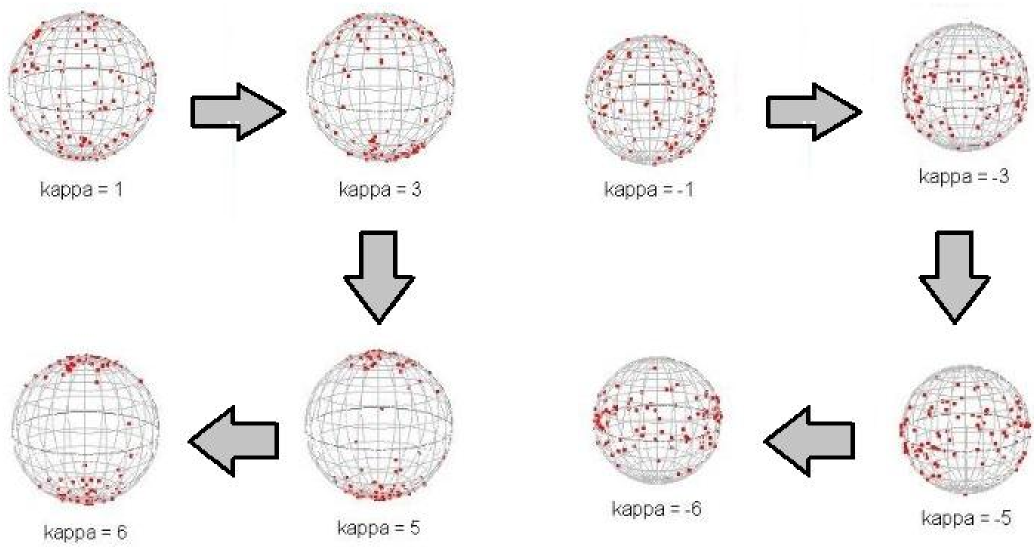


Figure 4 – Synthetic data from the W distribution over the unit sphere for positive and negative values of κ.

theory. Additionally, the best results were related to the most concentrated ($\kappa = 6$) events, concordantly with the Figure 4.

In all cases the bias-corrected estimator outperformed the standard MLE, mainly for the small values of κ. For instance, for $n = 20$ and $\kappa = 1$, the pair (bias, MSE) of MLE reduced from $(-1.084, 5.5753)$ to $(-0.6044, 2.7702)$, using the proposed bias expression.

Table 1 – Values for bias and MSE of standard and corrected ML estimates under the bipolar cases.

κ	n	$\hat{\kappa}$ Bias	$\hat{\kappa}$ MSE	$\tilde{\kappa}$ Bias	$\tilde{\kappa}$ MSE
1	20	1.083767	5.575265	−0.604372	2.770207
	30	0.839706	3.633523	−0.538646	2.220319
	50	0.589435	2.208639	−0.491987	1.544695
3	20	0.279182	4.582195	−0.092837	2.012087
	30	0.185040	2.689280	−0.057160	0.798357
	50	0.090915	0.462824	−0.036790	0.329417
5	20	0.225652	1.240230	−0.047078	0.701402
	30	0.145837	0.761402	−0.024675	0.442688
	50	0.054362	0.366539	−0.012063	0.273039
6	20	0.202915	1.184747	−0.025526	0.652320
	30	0.121577	0.678365	−0.011676	0.371082
	50	0.043542	0.306849	−0.001939	0.187414

The values of MSE and bias are shown in Table 2. For this case, we expect that as we reduce the value of κ the values of the pairs (bias, MSE) decrease, which in fact occurs. For example, for $\kappa = -1$ and sample size $n = 20$, we have that the pair (bias, MSE) decreases of $(0.482635, 5.270421)$ to $(0.269154, 3.739063)$, which corresponds to a reduction of almost 50%. The proposed bias expressions deliver good results when the sample are small and the values of the concentration parameter κ are larger.

Table 2 – Values for bias and MSE of standard and corrected ML estimates under the girdle cases.

κ	n	$\hat{\kappa}$ Bias	$\hat{\kappa}$ MSE	$\tilde{\kappa}$ Bias	$\tilde{\kappa}$ MSE
−1	20	−0.482635	5.270421	0.269154	3.739063
	30	−0.461038	3.029898	0.186358	2.852630
	50	−0.408045	1.954804	0.138775	1.447265
−3	20	−0.261471	4.650727	0.101254	2.115039
	30	−0.205408	2.927740	0.056094	0.824375
	50	−0.102762	0.498553	0.035949	0.341519
−5	20	−0.226860	1.267251	0.061973	0.820411
	30	−0.160697	0.791517	0.045966	0.481658
	50	−0.065887	0.402002	0.017992	0.290492
−6	20	−0.199696	1.164198	0.030743	0.686374
	30	−0.101212	0.647820	0.026262	0.385643
	50	−0.059538	0.281609	0.002294	0.196863

2.5 Conclusion

This chapter has presented a bias correction estimation method for the concentration parameter at the W model. It was made according to the Cox-Snell methodology. We have proposed closed-form expressions the Fisher information matrix and the second-order bias of ML estimator of concentration parameter. Results from a Monte Carlo study have indicated our proposal outperforms the classical ML estimator for both bipolar and girdle cases.

3 MATHEMATICAL METHODS BASED ON THE MINIMUM L^2 DISTANCE

3.1 Introduction

As already discussed, in terms of observations nature, Directional Statistics refers to the collection of methodologies to deal with points on the unit circle, sphere, or hypersphere. These types of data arise naturally in various science problems, such like Neuroscience Leong *et al.*, (1998), Biology Ferguson, (1967), and Geology Embleton *et al.*, (1998), among others.

Nonparametric kernel methods for estimating densities of spherical data have been studied in Hall *et al.* (1987) and Bai *et al.* (1988). *Kernel density estimation* or, in general, kernel smoothing methods is a classical topic in Nonparametric Statistics. First papers to tackle this issue (in Euclidean or linear perspective) have been proposed by Akaike (1954), Rosemblatt (1956), and Parzen (1962). Subsequently, extensions of the kernel density methodology have been brought for different contexts, as smoothers for more complex data (censorship, truncation, and dependence) or dynamical models Muller, (2006). Some comprehensive mathematical treatments have been addressed in books by Silverman (1986), He (1992), and Ko *et al.* (1993), among others. Beyond the linear case (or Euclidean), the kernel density estimation has been also adapted to directional data. Hall *et al.* (1987) have defined two types of kernel estimators and given asymptotic formula for bias, variance, and square loss.

Other paramount concept in Statistics is the *robustness*. This notion is pretty wide, but a common motivation for all robustness branches seems to be minor sensibility to the presence of outliers. The development of robust procedures for directional data has gained a lot of attention over the last years. The fact that the variables belong to a compact set requires tailored proposals in ways that are different of those due to the Euclidean space. A survey about robust methods for circular data can be found in He (1992). Ko *et al.* (1993) have extended the classic M-estimators for location and concentration parameters in distributed von Mises data.

A manner of combining both nonparametric (kernel) estimators and robust methods is by means of the statistics based on *minimum distances*, what is called *statistical distances* (SDs). There are at least two possible implications for the use of SDs in statistical analysis. Firstly, SDs have been employed as minimum distance estimation. This class of estimators has highlighted itself by presenting as robust properties as asymptotic efficiency, see Beran (1977), Simpson (1989), and Lindsay (1994). Beran (1977) pioneered a robust estimation technique, named *minimum Hellinger distance estimation.* The latter is a result of minimizing a discrepancy between a parametric density and one of its suitable nonparametric estimators. Secondly, SDs have also been used as Goodness-of-Fit (GoF) tests. In this case, the closeness between the empirical distribution and assumed model is quantified through SDs.

This chapter aims to study a natural extension of the minimum distance estimators discussed by Cao *et al.* (1994). More precisely, under the assumption of distributed Watson directional nature, we derive a mathematical background to produce hypothesis and point statistical inference procedures as well as GoF techniques.

This chapter is organized as follows. In Section 3.2, a kernel density estimation for directional data is discussed and illustrated. Section 3.3 approaches the minimum distance estimation. Section 3.4 exhibits the main results of this chapter. In Section 3.5, some numerical results about SDs are presented. The main conclusions listed in Section 3.6 ends this chapter.

3.2 Kernel estimator for directional data

According to previous discussion, kernel density estimators has been adapted to different non-Euclidean spaces, such like directional data or a q-dimensional sphere, having as special spaces, the circle ($q = 1$) and the sphere ($q = 2$). Let $\boldsymbol{x}$ be a directional random vector having density $f(\boldsymbol{x})$. The support of $\boldsymbol{x}$ is denoted by

$$\boldsymbol{\Omega}_q = \Big\{ \boldsymbol{x} \in \mathbb{R}^{q+1} \,:\, x_1^2 + \cdots + x_{q+1}^2 = 1 \Big\} = \Big\{ \boldsymbol{x} \in \mathbb{R}^{q+1} \,:\, ||\boldsymbol{x}||^2 = 1 \Big\},$$

where $\|.\|$ represents the Euclidean norm. The Lebesgue measure over Ω_q should then satisfy

$$\int_{\Omega_q} f(\boldsymbol{x})\, \omega_q(\mathrm{d}\boldsymbol{x}) = 1.$$

Remark 1. *When there is no possible misunderstanding, ω_q will also denote the surface area of Ω_q:*

$$\omega_q = \omega_q(\Omega_q) = \frac{2\pi^{\frac{q+1}{2}}}{\Gamma(\frac{q+1}{2})}, \quad \text{for } q \geq 1,$$

where $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$ represents the Gamma function.

According to Hall *et al.* (1987) and Bai *et al.* (1988), a directional kernel density estimator can be formulated as follows. Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ be a n-points random sample over the unity sphere, a directional kernel density estimator for $f(\boldsymbol{x})$ is given by

$$\widehat{f}_n(\boldsymbol{x}) = \frac{C_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \boldsymbol{x}^\top \boldsymbol{x}_i}{h^2}\right), \quad \text{for} \quad \boldsymbol{x} \in \Omega_q, \tag{3.1}$$

where $^\top$ is the transpose operator, $L(\cdot)$ is the directional kernel, $h \geq 0$ is the bandwidth parameter, and $C_{h,q}(L)$ is a normalizing constant depending on the form of kernel $L(\cdot)$, the bandwidth h, and the dimension q. The inverse of the normalizing constant at $\boldsymbol{x} \in \boldsymbol{\Omega}_q$ is given by (Bai *et al.*, 1988)

$$c_{h,q}(L)^{-1} \;=\; \int_{\boldsymbol{\Omega}} L\left(\frac{1-\boldsymbol{x}^\top \boldsymbol{y}}{h^2}\right) w_q(\mathrm{d}\boldsymbol{y}) \;=\; h^q\,\lambda_{h,q}(L),$$

where $\lambda_{h,q}(L) = w_{q-1}\int_0^{2\,h^{-2}} L(r)\, r^{\frac{q}{2}-1}(2 - r\,h^2)^{\frac{q}{2}-1}\mathrm{d}r$. Figure 5 illustrates the used kernel function for a set of generated directional data.

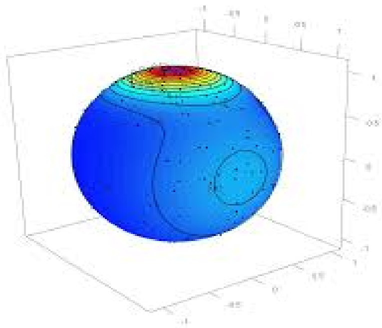


Figure 5 – Contour plot of the mixture of von Mises densities.

Various properties of the directional kernel density estimator (3.1) have been proposed by Bai *et al.* (1988), e.g., its $\mathbb{L}_1$-norm consistency. Additionally, a central limit theorem for the integrated squared error of the estimator as well as the expression for the bias have been derived by Zhao *et al.* (2001), under the following regularity conditions:

1. Extend $f(\cdot)$ from $\boldsymbol{\Omega}_q$ to $\mathbb{R}^{q+1}\backslash\{\mathbf{0}\}$ by defining $f(\boldsymbol{x}) \equiv f(\boldsymbol{x}\backslash\|\boldsymbol{x}\|)$ for all $\boldsymbol{x} \neq 0$. Assume (i) the gradient vector $\nabla\left[f(\boldsymbol{x})\right] = \left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_1}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_{q+1}}\right)^\top$ and the associated Hessian matrix $\mathbb{H}\left[f(\boldsymbol{x})\right] = \left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\, x_j}\right)_{1\leq i,j\leq q+1}$ exist and (ii) they are also continuous on $\mathbb{R}^{q+1}\backslash\{\mathbf{0}\}$ and integrable square.

2. Assume that the kernel function $L : [0,\infty) \to [0,\infty)$ is a bounded and Riemann integrable function such that

$$0 < \int_0^\infty L^k(r) r^{\frac{q}{2}-1} dr < \infty, \;\; for\;\; all\;\; q \geq 1, \qquad for \qquad k = 1, 2.$$

3. Further, consider that $\{h_n\,;\, n = 1, 2, \ldots\}$ is a sequence of positive numbers such that $h_n \to 0$ and $nh_n^q \to \infty$ as $n \to \infty$.

About the kernel function, have worked with

$$L(t) \;=\; 1.5\left(1 \;-\; t^2\right)\mathbb{I}(0 < t < 1).$$

The choice of the smoothing parameter is one of major problems in density estimation. In the linear case, several authors dealt with the problem of providing an automatic procedure to select the bandwidth. For directional data, a bandwidth was considered in Hall *et al.* (1987), while a bandwidth selector was considered in Taylor(2008). In this research, we choose the bandwidth $h \in [0.2, 1.4]$.

3.3 Minimum distance estimators

Let $\boldsymbol{\Omega}_q \subset \mathbb{R}^{q+1}$ be the q-dimensional unit sphere of radius one centered at $\mathbf{0}$. Let $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ be observed sample drawn from a random vector $\boldsymbol{x}$ with density $f(\cdot)$ over the support $\boldsymbol{\Omega}_q$. Let $f_\theta(\boldsymbol{x})$ be a family of density functions parametrized with vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q$. Consider then the problem of estimating $\boldsymbol{\theta}$.

Given a distance function in the space of density functions, we define a *minimum distance-based estimator* θ_n as any value satisfying

$$D(\hat{f}_n, f_{\boldsymbol{\theta}_n}) \leq \inf_{\boldsymbol{\theta}} D(\hat{f}_n, f_{\boldsymbol{\theta}}) + \delta_n, \tag{3.2}$$

where $\delta_n \to 0$ as $n \to \infty$ and $\hat{f}_n$ is a nonparametric density estimator. Notice that if the infimum of $D(\hat{f}_n, f_\theta)$ is achieved, then the estimator can be defined as follows:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} D(\hat{f}_n, f_\theta). \tag{3.3}$$

Some standard choices for D are the distances induced from the L^p norm or from the L^∞ norm, i.e.,

$$D_p(\hat{f}_{hn}, f_\theta) = \int \left\| \hat{f}_{hn}(\boldsymbol{x}) - f_\theta(\boldsymbol{x}) \right\|^p \omega_d(d\boldsymbol{x}) \text{ and } D_\infty(\hat{f}_{hn}, f_\theta) = sup_{\boldsymbol{x}} \left\| \hat{f}_{hn}(\boldsymbol{x}) - f_\theta(\boldsymbol{x}) \right\|^p \omega_d(d\boldsymbol{x}),$$

respectively. The Hellinger distance

$$D_H(\hat{f}_{hn}, f_\theta) = \int (\sqrt{\hat{f}_{hn}(\boldsymbol{x})} - \sqrt{f_\theta(\boldsymbol{x})})^2 \omega_d(d\boldsymbol{x})$$

could be another choice. This distance was studied by Beran (1977), who showed that robustness and efficiency properties could be obtained using this distance. Another alternative could be to consider D as discrepancy measures or divergences. This approach was considered by Agostinelli (2007) for circular data and by Basu *et al.* (1994) for Euclidean data.

Garcia *et al.* (2013) introduced alternatives. One of them assumes that the underlying distribution is von Mises and the other considers a mixture of von Mises densities. In this Chapter we give a mathematical background for extension of that problem. We assume the real Watson distribution. However, as Cao *et al.* (1995) remarked, it seems reasonable to take into account the particular features of the problem and consider another alternative.

We consider an automatic bandwidth as in Cao *et al.* (1995). The idea is to incorporate the smoothing parameter h_n as an additional component in the vector $\boldsymbol{\theta}$. Therefore, we can obtain simultaneously an automatic bandwidth and the estimator of $\boldsymbol{\theta}$ as follows:

$$(\tilde{\boldsymbol{\theta}}, \tilde{h}_n) = \arg \min_{\boldsymbol{\theta}, \boldsymbol{h}} D(\widehat{f}_n, f_\theta). \tag{3.4}$$

The proposed estimators, defined (3.4), are strongly consistent and asymptotically normally distributed. These properties follow easily using analogous arguments to those considered by Cao *et al.* (1995). The following assumptions are needed in order to obtain the desired results.

Assumptions:

- A The nonparametric estimator should satisfy $\lim_{n\to\infty} D(\hat{f}_{h_n}, f_{\boldsymbol{\theta}}) = 0$ as $f_{\boldsymbol{\theta}}$.

- B For all $\boldsymbol{\theta}_0 \in \Theta$ and a sequence $\boldsymbol{\theta}_m \subset \boldsymbol{\Theta}$ such that $\lim_{m\to\infty} D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_m}) = D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})$; we have that $\lim_{m\to\infty} \boldsymbol{\theta}_m = \boldsymbol{\theta}_0$.

- C The Kernel $L : [0, \infty) \to [0, \infty)$ is a bounded and integrable function with compact support. For the following hypotheses, we are considering the extension of f in $\mathbb{R}^{d+1}\backslash\{\mathbf{0}\}$ given by $f(\boldsymbol{x}\backslash \|\boldsymbol{x}\|)$ for all $\boldsymbol{x} \neq 0$, where $\|\boldsymbol{x}\|$ denotes the Euclidean norm of $\boldsymbol{x}$.

- D The density function $f_{\boldsymbol{\theta}}$ is such that (for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$)

  - D1 $\nabla[f(\boldsymbol{x})]$ and the Hessian matrix $\mathbb{H}[f(\boldsymbol{x})]$ exist and are continuous in $\mathbb{R}^{d+1}\backslash\{\mathbf{0}\}$.
  - D2 $\frac{\partial f_\theta(\boldsymbol{x})}{\partial \theta} = \left(\frac{\partial f_\theta(\boldsymbol{x})}{\partial \theta_1}, \cdots, \frac{\partial f_\theta(\boldsymbol{x})}{\partial \theta_q}\right)^\top$ is integrable with respect to the measure generated from $L(\cdot)$ and $f_{\boldsymbol{\theta}}$.

- E $\Psi(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} - \mathbb{E}_{\boldsymbol{\theta}}(\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}})$ is differentiable function with respect to $\boldsymbol{\theta}$.

- F If we denote by $\boldsymbol{\theta}_0$ the true value of the parameter, the matrix $\boldsymbol{A} = \mathbb{E}_{\boldsymbol{\theta}_0}(\frac{\partial \Psi(\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is non singular.

- G The sequences $nh_n^4 \to 0$ as $n \to \infty$ and $h_n^4 c(h_n) \to \lambda$ as $n \to \infty$ with

$$\lambda^{-1} = 2^{(d/2)-1} \tau_{d-1} \int_0^\infty L(r)\, r^{(d/2)-1}\, \mathrm{d}r,$$

  where τ_d is the area of S_d and $\tau_{d-1} = 2\pi^{d/2}\Gamma(d/2)$ for $d \geq 1$.

The asymptotic behavior of estimators (3.4) is determined by the follow result.

Theorem 3.1. *Under assumptions C to G, the estimator defined in* (3.3) *satisfies*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1})$$

where Ψ *and* $\boldsymbol{A}$ *were defined in conditions* E *and* F*, respectively and* $\Sigma = \mathbb{E}_{\theta_0}(\Psi(\boldsymbol{x}, \theta_0)\Psi(\boldsymbol{x}, \theta_0)^\top)$.

3.4 The proposed distance for the Watson distribution

Now we are in position to derive the theoretical contributions of this chapter. Consider $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote a observed sample which is obtained from the random vector following the real Watson model, $\boldsymbol{x} \sim W(\kappa, \boldsymbol{\mu})$ with pdf

$$f_W(\boldsymbol{x}) \;=\; c_1^{-1}(\kappa)\, \exp\left\{ \kappa\, (\boldsymbol{x}^\top \boldsymbol{x})^2 \right\},$$

where $c_1^{-1}(\kappa)$ is the normalizing constant defined previously, κ is the concentration parameter and $\boldsymbol{\mu}$ is the mean axis. We opt by studying the expression of the squared distance

$$D_2(\widehat{f}_{h_n}, f_W) \;=\; \int_{\boldsymbol{\Omega}_q} ||\widehat{f}_{h_n}(\boldsymbol{x}) \;-\; f_W(\boldsymbol{x})||^2\, w_q(\mathrm{d}\boldsymbol{x}),$$

where, taking $c(h_n) = c_{h,q}(L)$, after some algebra,

$$\widehat{f}_{h_n}(\boldsymbol{x}) \;=\; \frac{1.5\, c(h_n)}{n\, h_n^4} \left\{ n\,(h_n^4 - 1) \;+\; \sum_{i=1}^{n} \left[2\, \boldsymbol{x}^\top \boldsymbol{x}_i \;-\; (\boldsymbol{x}^\top \boldsymbol{x}_i)^2 \right] \right\}.$$

The following theorem furnishes the distance expression we use.

Theorem 3.2. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be an observed sample from $\boldsymbol{x} \sim W(\kappa, \boldsymbol{\mu})$ with density $f_W(\boldsymbol{x})$ and $\widehat{f}_{h_n}$ a (according to Epanechnikov) kernel estimator for its density, then the expression holds:* $D_2(\widehat{f}_{h_n}, f_W) \;=\; \frac{4\,\pi}{3}\, c^2(h_n) \left(\frac{h_n^4 - 1}{h_n^4}\right) \;-\; \frac{2}{n}\, c^2(h_n) \left(\frac{h_n^4 - 1}{h_n^8}\right) \sum_{i=1}^n \boldsymbol{x}_i^\top \begin{bmatrix} \frac{4\,\pi}{15} & 0 & 0 \\ 0 & \frac{4\,\pi}{15} & 0 \\ 0 & 0 & \frac{2}{15} \end{bmatrix} \boldsymbol{x}_i$

$$+ \left(\frac{9\, c^2(h_n)}{h_n^8}\right) \left\{ \bar{\boldsymbol{x}}^\top \begin{bmatrix} \frac{4\,\pi}{15} & 0 & 0 \\ 0 & \frac{4\,\pi}{15} & 0 \\ 0 & 0 & \frac{2}{15} \end{bmatrix} \bar{\boldsymbol{x}} \;-\; \underbrace{\int_{\boldsymbol{\Omega}_3} \left(\bar{\boldsymbol{x}}^\top \boldsymbol{x}\right) \left(\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x}\right)\, w_3(\mathrm{d}\boldsymbol{x})}_{A} \right\}$$

$$+ \left(\frac{9\, c^2(h_n)}{4\, h_n^8}\right) \underbrace{\int_{\boldsymbol{\Omega}_3} \left(\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x}\right)^2\, w_3(\mathrm{d}\boldsymbol{x})}_{B} \;-\; 3\, c(h_n) \left(\frac{h_n^4 - 1}{h_n^4}\right) \;+\; \frac{c_1(2\,\kappa)}{c_1^2(\kappa)}$$

$+ \left(\frac{3\, c(h_n)}{n\, h_n^4}\right) \sum_{i=1}^n \boldsymbol{x}_i\, \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top)\, \boldsymbol{x}_i$, *where* $\bar{\boldsymbol{x}} \;=\; n^{-1} \sum_{i=1}^n \boldsymbol{x}_i$, $\boldsymbol{S} \;=\; n^{-1} \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^\top$, *from Kurz et. al. (2016)*

$$\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) \;=\; \boldsymbol{M}\, \mathrm{diag}\left(D_d(\kappa), \frac{1 - D_d(\kappa)}{d - 1}, \cdots, \frac{1 - D_d(\kappa)}{d - 1} \right) \boldsymbol{M}^\top,$$

$\boldsymbol{M}$ *is an arbitrary rotation matrix whose first column is* $\boldsymbol{\mu}$,

$D_q(\kappa) = M(3/2,(d+1)/2,\kappa)[q\,M(3/2,d/2,\kappa)]^{-1}$, $A = \int_0^1 \int_0^{2\pi} \int_0^{\pi}[\bar{x}_1\, r\, \sin(\theta)\, \cos(\phi) + \bar{x}_2\, r\, \sin(\theta)\, \sin(\phi) + \bar{x}_3\, r\, \cos(\theta)]$
$\times\ [\, s_{11}\, r^2\, \sin^2(\theta)\, \cos^2(\phi) + s_{22}\, r^2\, \sin^2(\theta)\, \sin^2(\phi) + s_{33}\, r^2\, \cos^2(\theta)$
$+\ 2\, s_{12}\, r^2\, \sin^2(\theta)\, \cos(\phi)\, \sin(\phi) + 2\, s_{13}\, r^2\, \sin(\theta)\, \cos(\theta)\, \cos(\phi)$
$+\ 2\, s_{23}\, r^2\, \sin(\theta)\, \cos(\theta)\, \sin(\phi)] \times r^2\, \sin(\theta) \qquad \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r$
and $B = \int_0^1 \int_0^{2\pi} \int_0^{\pi}\ [\, s_{11}\, r^2\, \sin^2(\theta)\, \cos^2(\phi) + s_{22}\, r^2\, \sin^2(\theta)\, \sin^2(\phi) + s_{33}\, r^2\, \cos^2(\theta)$
$+\ 2\, s_{12}\, r^2\, \sin^2(\theta)\, \cos(\phi)\, \sin(\phi) + 2\, s_{13}\, r^2\, \sin(\theta)\, \cos(\theta)\, \cos(\phi)$
$+\ 2\, s_{23}\, r^2\, \sin(\theta)\, \cos(\theta)\, \sin(\phi)]^2 \times r^2\, \sin(\theta) \qquad \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r.$

Moreover, the distance $D_2(\widehat{f}_{h_n}, f_W)$ should satisfy the condition

$$0 = \frac{\partial\, D_2(\widehat{f}_{h_n}, f_W)}{\partial\, \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}=(\widehat{\kappa},\widehat{\boldsymbol{\mu}})} = \left[\int_{\boldsymbol{\Omega}_q} \widehat{f}_{h_n}(\boldsymbol{x})\, \frac{\partial f_W(\boldsymbol{x})}{\partial \boldsymbol{\theta}}\, w_q(\mathrm{d}\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{\theta}}\left(\frac{\partial f_W(\boldsymbol{x})}{\partial \boldsymbol{\theta}}\right)\right]\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$
$$= \frac{1}{n}\sum_{i=1}^{n} \Psi_h(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}}),$$

where expressions for $\Psi_h(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}})$ are given by the following theorem.

Corollary 3.1. *The expression for* $\Psi_h(\boldsymbol{x}_i; \boldsymbol{\theta})$ *(3.5) if*

(i) in terms of κ, $\Psi_h(\boldsymbol{x}_i; \boldsymbol{\theta}) = \left(\frac{1.5\, c(h_n)\, c_3(\kappa)}{h_n^4}\right) \boldsymbol{x}_i^\top\, \mathbb{E}_{(\kappa,\boldsymbol{\mu})}\left(\boldsymbol{x}\boldsymbol{x}^\top\right) \boldsymbol{x}_i$
$-\ \left(\frac{1.5\, c(h_n)}{h_n^4}\right) \boldsymbol{x}_i^\top \left[\sum_{j=1}^3 \sum_{k=1}^3 \mu_j\, \mu_k\, \mathbb{E}_{(\kappa,\boldsymbol{\mu})}\left(X_j\, X_k \boldsymbol{x}\boldsymbol{x}^\top\right)\right] \boldsymbol{x}_i$
$-\ n\, \frac{c^2(\kappa)}{c(2\,\kappa)} \left[\boldsymbol{\mu}^\top\, \mathbb{E}_{(2\,\kappa,\boldsymbol{\mu})}\left(\boldsymbol{x}\boldsymbol{x}^\top\right) \boldsymbol{\mu} - c_3(\kappa)\right]$,
where $c_3(\kappa) = M(3/2,(q+2)/2,\kappa)\, [q\, M(1/2, q/2,\kappa)]^{-1}$, *and*

(ii) in terms of $\boldsymbol{\mu}_k$ *for* $k{-}1,2,3$, $\Psi_h(\boldsymbol{x}_i; \boldsymbol{\theta}) = -\left(\frac{3\, c(h_n)\, \kappa}{h_n^4}\right) \boldsymbol{x}_i^\top \left[\sum_{j=1}^3 \mu_j\, \mathbb{E}_{(\kappa,\boldsymbol{\mu})}\left(X_j\, X_k \boldsymbol{x}\boldsymbol{x}^\top\right)\right] \boldsymbol{x}_i$
$-\ \left(\frac{2\, n\, c^2(\kappa)\, \kappa}{c(2\,\kappa)}\right) \boldsymbol{\mu}^\top\, \mathbb{E}_{(2\,\kappa,\boldsymbol{\mu})}\left(\boldsymbol{x}\, \boldsymbol{x}^\top\right) \boldsymbol{e}_k$,
where $\boldsymbol{e}_k$ *is a null vector with exception of the* k*th element, which is* 1. *The terms derived from this corollary are in Appendix C.*

3.5 Descriptive experiments for the validation of $D_2(\widehat{f}_{h_n}, f_W)$

This section is an initial discussion about the potentiality of the use of $D_2(\widehat{f}_{h_n}, f_W)$ in statistical analysis. Note that there are several factors we need to control in order to apply this tool, e.g., length of bandwidth, estimators to be used and their asymptotic behavior, sample size, concavity of estimation or GoF criteria, and geometrical properties of data (for data obtained from statistical manifolds), among others. Here, beyond its canonical conditions (consequent of definition, *Nonegativity*, *Symmetry*, and *Idempotent*), we want to verify (for $\boldsymbol{\theta}_i = (\kappa_i, \boldsymbol{\mu}_i)$)

$$D_2(\widehat{f}_{h_n}(\boldsymbol{x}), f_W(\boldsymbol{x}, \boldsymbol{\theta}_1)) \leq D_2(\widehat{f}_{h_n}(\boldsymbol{x}), f_W(\boldsymbol{x}, \boldsymbol{\theta}_2)) \tag{3.5}$$

when the n-points observed sample is obtained from $\boldsymbol{x} \sim W(\kappa_1, \boldsymbol{\mu}_1)$. As discussed, the parameter κ indicates concentration and, therefore, if one changes the value of κ kipping $\boldsymbol{\mu}$ fixed is intuitive that distance does not change once this phenomenon is represented by two overlapped axes centralized two points clouds with different dispersion. Thus, we admit that $\kappa_1 = \kappa_2 = \kappa$ and $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ in (3.5).

We developed a Monte Carlo simulation framework with one thousand replicas in order to quantify the performance of the developed SD. To this end, we adopt:

(i) $\kappa = 1, 5$, and 15, represent scenarios from high to low degree of dispersion.

(ii) $\boldsymbol{\mu}_i = [\sin(\theta_i)\ \cos(\phi_i), \sin(\theta_i)\ \sin(\phi_i), \cos(\theta_i)]$ for $\theta_i = \phi_i = i = 0.1, 0.5, 1, 1.5, 2$, where $\boldsymbol{\mu}_3$ is considered the real data reference and the minimum arc lengths over sphere between $\boldsymbol{\mu}_3$ and $\boldsymbol{\mu}_i$, $d_g(\boldsymbol{\mu}_3, \boldsymbol{\mu}_i) = \arccos(|\boldsymbol{\mu}_3^\top \boldsymbol{\mu}_i|)$, are

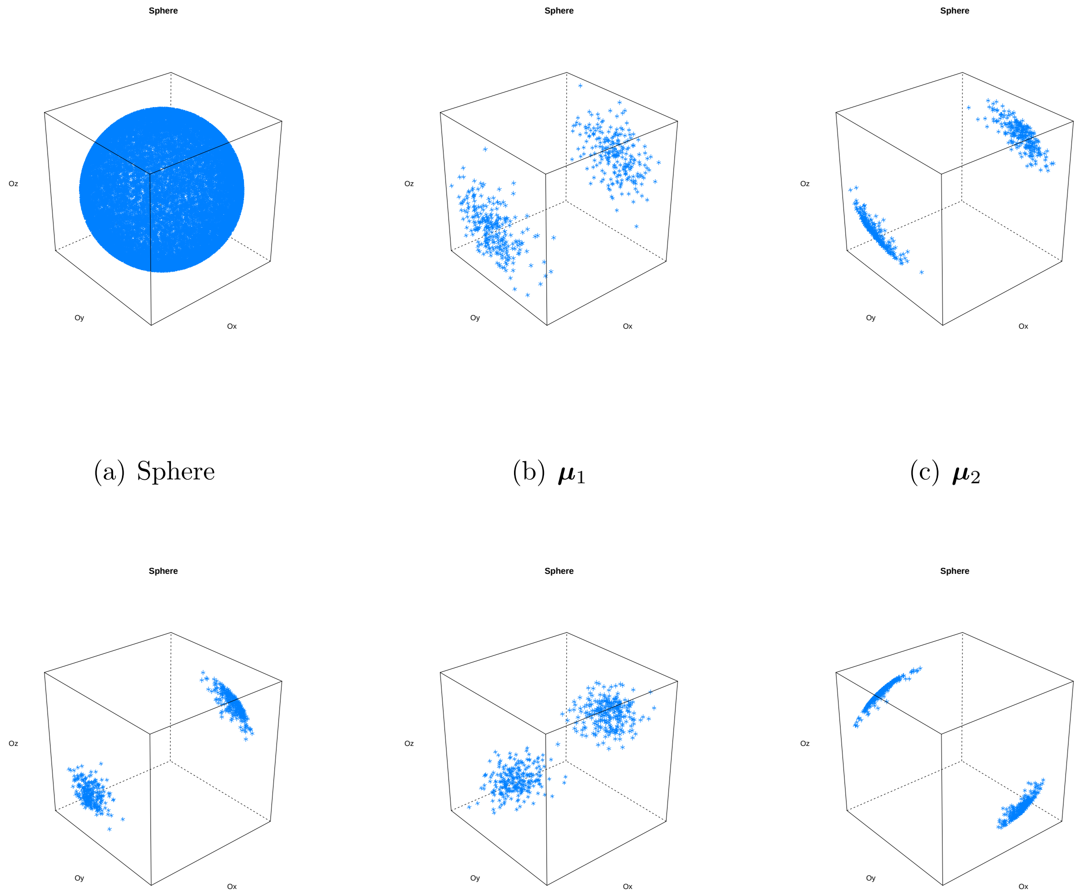
Index, i	1	2	3	4	5
$d_g(\boldsymbol{\mu}_3, \boldsymbol{\mu}_i)$	0.940	0.595	0.000	0.684	1.381

.

Figure 6 illustrates the cases we adopt for $\kappa = 25$ (which is known as almost bipolar phenomena). It seems intuitive the smallest arc length for $(\boldsymbol{\mu}_2, \boldsymbol{\mu}_3)$ and the highest for $(\boldsymbol{\mu}_3, \boldsymbol{\mu}_5)$.

(iii) As, from pilot study, there was not great changes under variation of sample sizes, we use $n = 200$.

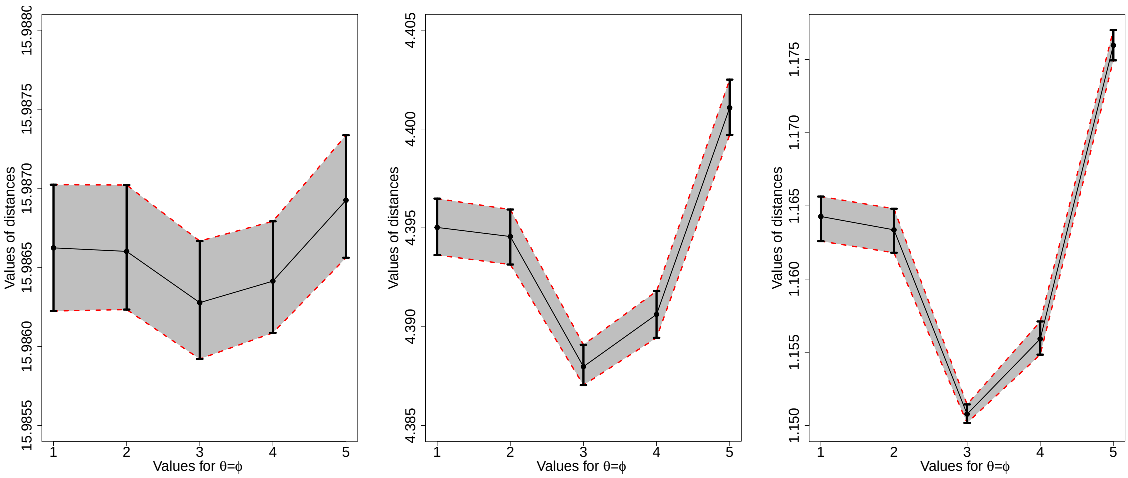
Figure 7 presents numerical results of values of the median, minimum, maximum of SDs. With respect to the influence of κ, one has that the increasing at value of κ implies more accurate performances, i.e., (i) the descriptive measures at the thruth scenarios, $D_2(\widehat{f}_{h_n}, f_W(\boldsymbol{x}, \kappa, \boldsymbol{\mu}_3))$, are more disjoint of the remainder ones, (ii) the variability measures (such like amplitude) at the thruth scenarios tend to be smaller, and (iii) the values of the SD are closer of zero. On the other hand, in terms of $\boldsymbol{\mu}_i$, it is noticeable $D_2(\widehat{f}_{h_n}, f_W(\boldsymbol{x}, \kappa, \boldsymbol{\mu}_3)) < D_2(\widehat{f}_{h_n}, f_W(\boldsymbol{x}, \kappa, \boldsymbol{\mu}_1)) < D_2(\widehat{f}_{h_n}, f_W(\boldsymbol{x}, \kappa, \boldsymbol{\mu}_5))$ in concordance with the values of arc lengths, but the effect of κ can affect distinctions of scenarios whose rotation deference is low, as happened with spherical samples indexed by $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_4$.



(a) Sphere (b) $\boldsymbol{\mu}_1$ (c) $\boldsymbol{\mu}_2$

(d) $\boldsymbol{\mu}_3$ (e) $\boldsymbol{\mu}_4$ (f) $\boldsymbol{\mu}_5$

Figure 6 – Illustration of simulation cases for $\boldsymbol{\mu}_1^\top = [0.099334665, 0.009966711, 0.995004165]$, $\boldsymbol{\mu}_2^\top = [0.4207355, 0.2298488, 0.8775826]$, $\boldsymbol{\mu}_3^\top = [0.4546487, 0.7080734, 0.5403023]$, $\boldsymbol{\mu}_4^\top = [0.0705600, 0.9949962, 0.0707372]$, and $\boldsymbol{\mu}_5^\top = [-0.3784012, 0.8268218, -0.4161468]$



(a) $D_2(\widehat{f}_{h_n}, f_W)|\kappa = 1$ (b) $D_2(\widehat{f}_{h_n}, f_W)|\kappa = 5$ (c) $D_2(\widehat{f}_{h_n}, f_W)|\kappa = 15$

Figure 7 – Performance of statistical distances, under the lexicographic relation: 1 - data $\sim W(\kappa, \mu_3)$ vs. model $\sim W(\kappa, \mu_1)$, 2 - data $\sim W(\kappa, \mu_3)$ vs. model $\sim W(\kappa, \mu_2)$, 3 - data $\sim W(\kappa, \mu_3)$ vs. model $\sim W(\kappa, \mu_3)$, 4 - data $\sim W(\kappa, \mu_3)$ vs. model $\sim W(\kappa, \mu_4)$, and 5 - data $\sim W(\kappa, \mu_3)$ vs. model $\sim W(\kappa, \mu_5)$.

3.6 Conclusion

This chapter has addressed the derivation of the quadratic statistical distance for directional data; in particular, following the real Watson distribution. In parallel, various theoretical properties for this model have been derived and they can be used in other contexts, such like Asymptotic Theory and Multivariate Analysis. Even though the numerical part of this chapter is an initial discussion, results pointed out the use of statistical distances as inference or GoF criteria can be a good alternative in directional data analysis.

4 BOOTSTRAP AND PERMUTATION TESTS FOR AXIAL DATA

4.1 Introduction

The statistical analysis of directional data, represented by points on the surface of the unit sphere in $\mathbb{R}^q$, denoted by $\mathbb{S}_{q-1} = x \in \mathbb{R}^q : \boldsymbol{x}^\top \boldsymbol{x} = 1$ was widely developed by Watson (1983), Fisher et al. (1987), Fisher (1993), Mardia and Jupp (2000), among other authors. With recent technological advances, the statistical analysis on the sphere has bee requested in many fields see Kent (1994), Bingham (1974), Watson (1965) and Fisher, Lewis, and Embleton (1987).

The sophistication of data mining resources has allowed to study phenomena previously not explored; however, many of them require specialized probabilistic models and statistical methods for analyzing resulting data. In particular, this is the case when one wishes to analyze spherical data, as discussed subsequently.

Spherical data can be easily represented as a straight line on normed space, see Fisher, Lewis, and Embleton (1987). In this case resulting data are classified as directional, if such line is directed, and axial in otherwise. Directional data can be represented as points on a sphere; while, axial data represent pairs of antipodal points (i.e., opposite points) on a sphere. The Watson model is often used for describing axial data, especially those defined on the circumference and sphere supports see Watson (1965) and Fisher, Lewis, and Embleton (1987).

Directional data arise in many scientific areas, such as biology, geology, machine learning, text mining, bioinformatics, among others. An important problem in directional statistics and shape analysis, as well in other areas of statistics, is to test the null hypothesis of a common mean vector or polar axis across several populations. This problem was already treated for circular data and spherical data by several authors, such as Stephens (1969), Underwood and Chapman (1985), Anderson and Wu (1995), Harrison et al. (1986), Jammalamadaka and SenGupta (2001), among others. However, there has been relatively little discussion of nonparametric bootstrap approaches to this problem.

In this chapter, we assume that such data are well described by the Watson distribution.This model is equipped by two parameters: dominant axis and concentration. The last parameter is easily interpretable because it is related to both position and dispersion of axes over unit sphere or circle for three and bi-dimensional data, respectively. Many articles employed the Watson distribution as probabilistic assumption. Li and Wong (1993) proposed a random number generator for the Watson distribution based on the acceptance-rejection method.

To detect outlier observations in Watson distributed samples, Figueiredo (2007) provided some important hypothesis tests in terms of the likelihood ratio. Recently, Sra and Karp (2013) presented some theoretical aspects of the Watson model. In particular, asymptotic behavior for the maximum likelihood estimators were investigated.

In this chapter, we propose two-samples divergence-based hypothesis tests involving the concentration parameter of the Watson distribution; i.e., two statistical procedures to identify if two axial samples are similarly concentrated. Subsequently, these expressions are redefined as hypothesis tests within the h-ϕ divergence class proposed by Salicru et al. (1994). By means of a Monte Carlo simulation study, we quantify the performance of the proposed methods. To that end, their empirical test sizes and powers are used as assessment criteria.

The remainder of this chapter is organized as follows. Hypothesis test based on stochastic distances is given Sec. 4.2. A short background about hypothesis tests based on statistical information theory measures using permutation tests and bootstrap is given in Sec. 4.3 In Sec. 4.4, numerical results involving the proposed methods are presented and a application to real data. Finally, main contributions are listed and briefly rediscussed in Sec. 4.5.

4.2 Hypothesis test based on stochastic distances

This section presents the background for providing new hypothesis tests for the concentration parameter κ. In what follows it is adopted that divergence measure, say d, means any non-negative function between two Watson probability densities which satisfies the definiteness property (for two densities $f(x)$ and $g(x)$, $f(x) = g(x)$ implies $d(f, g) = 0$). Moreover, a symmetrized divergence is understood as a distance measure Frery, Nascimento, and Cintra (2014).

Assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are two vectors equipped with densities $f(z, \theta_1)$ and $f(z, \theta_2)$, respectively, where $\theta_1 = (\kappa_1, \mu_1)^\top$ and $\theta_2 = (\kappa_2, \mu_2)^\top$ are parameter vectors. The densities are assumed to share a common support $\boldsymbol{S}^{p-1}$. The (h, ϕ)-divergence d_ϕ^h is defined by Pardo (2005)

$$d_\phi^h = h\left(\int_{\mathbb{S}_{p-1}} \phi\left(\frac{f(z;\theta_1)}{f(z;\theta_2)}\right) f(z;\theta_2)dz\right), \tag{4.1}$$

where $h : (0, \infty) \to [0, \infty)$ is a strictly increasing (or decreasing) function with $h(0) = 0$ and $\phi : (0, \infty) \to [0, \infty)$ is a convex (or concave) function such that $\phi(1) = 0$, $\phi(0/0) = 0$ and $\phi(x/0) = \lim_{x\to\infty} \phi(x)/x$ (Salicru et al. 1994, 374). Well-known divergences can be obtained from (4.1) through choices of h and ϕ; such as (i) Renyi, (ii) Bhattacharyya and (iii) Hellinger. In this chapter we work with divergence measures which applied to the Watson distribution specifically divergence distance Renyi given by

$$d_\phi^h = \frac{\beta}{\beta - 1} \log \int f_x^\beta f_y^{1-\beta}$$

Nascimento et al. (2018) proposed new hypothesis tests in order to verify if two Watson distributed samples are distinct. In particular, they emphasize the situation in which samples have different degrees of concentration around the same dominant axis. In addition, the following theorem and corollary were used for this discussion.

Theorem 4.1 (General support for calculating the new divergence)**.** *Let $\boldsymbol{x} \sim W_p(\mu_1, \kappa_1)$ and $W_p(\mu_2, \kappa_2)$ having densities $f(z; \mu_1, \kappa_1)$ and $f(z; \mu_2, \kappa_2)$, respectively, then the following identify holds (for $\beta \in \mathbb{R}_+ - 1$)*

$$\begin{aligned} \boldsymbol{K}_\beta(\kappa_1, \mu_1, \kappa_2, \mu_2) &\equiv \int_{\mathbb{S}^{p-1}} f^\beta(z; \mu_1, \kappa_1) f^{1-\beta}(z; \mu_2, \kappa_2) dz \\ &= \frac{(2\pi^{p/2})^{-1} \Gamma\left(\frac{p}{2}\right) C(\beta\kappa_1\mu_1{\mu_1}^\top + (1-\beta)\kappa_2\mu_2{\mu_2}^\top)}{M^\beta(\frac{1}{2}, \frac{p}{2}, \kappa_1) M^{1-\beta}(\frac{1}{2}, \frac{p}{2}, \kappa_2)}, \end{aligned}$$

where

$$C(\beta\kappa_1\mu_1{\mu_1}^\top + (1-\beta)\kappa_2\mu_2{\mu_2}^\top) \equiv \int_{\mathbb{S}^{p-1}} \exp\left\{\sum_{i=1}^p \lambda_i(\kappa_1, \mu_1, \kappa_2, \mu_2)\right\} dz$$

and $\lambda_i(\kappa_1, \mu_1, \kappa_2, \mu_2)$ is the ith eigenvalue of the matrix $(1-\beta)\kappa_2\mu_2{\mu_2}^\top - \beta\kappa_1\mu_1{\mu_1}^\top$. This theorem is derived in Appendix D. With respect the computation of $C(.)$, Kume and Wood (2005) have derived saddlepoint approximations to an expression of this kind. In future studies, other divergence-based hypothesis tests can be deduced from Theorem 4.1. The next corollary is obtained from the Theorem 4.1 by assuming $\mu_1 = \mu_2 = \mu$.

Corollary 4.1. *Let $x \sim W_p(\mu, \kappa_1)$ and $y \sim W_p(\mu, \kappa_2)$ having densities $f(z; \mu, \kappa_1)$ and $f(z; \mu, \kappa_2)$, respectively, then $K_\beta(\kappa_1, \mu, \kappa_2, \mu)$ collapses in the identify*

$$\boldsymbol{K}_\beta(\kappa_1, \kappa_2) = \frac{M(\frac{1}{2}, \frac{p}{2}, \beta\kappa_1 + (1-\beta)\kappa_2)}{M^\beta(\frac{1}{2}, \frac{p}{2}, \kappa_1) M^{1-\beta}(\frac{1}{2}, \frac{p}{2}, \kappa_2)} \to \frac{M(\frac{1}{2}, \frac{p}{2}, \frac{\kappa_1+\kappa_2}{2})}{\sqrt{M(\frac{1}{2}, \frac{p}{2}, \kappa_1) M(\frac{1}{2}, \frac{p}{2}, \kappa_2)}}.$$

An outline of the proof of this result is given in Appendix E. Now, consider the study of some properties of $K_\beta(.,.)$. As discussed before, $\kappa > 0$ tackles the bipolar case; while $\kappa < 0$ the girdle. Combined with the result $M(a, b; x) = \exp^x M(b-a, b; x)$, we have: Let $\kappa_1, \kappa_2 \leq 0$,

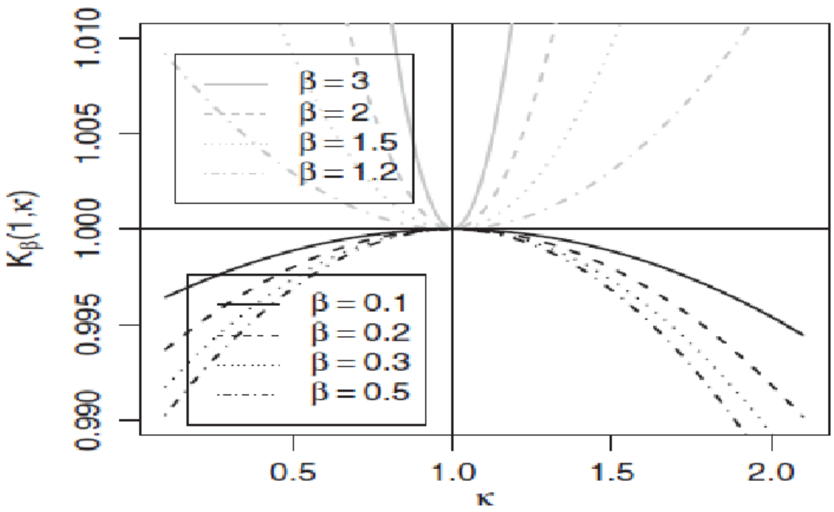


Figure 8 – Illustration of function $K_\beta(1, \kappa)$ for several values of β.

$$\boldsymbol{K}_\beta(\kappa_1, \kappa_2) = \frac{M(\frac{p-1}{2}, \frac{p}{2}, \beta|\kappa_1| + (1-\beta)|\kappa_2|)}{M^\beta(\frac{p-1}{2}, \frac{p}{2}, |\kappa_1|)M^{1-\beta}(\frac{p-1}{2}, \frac{p}{2}, |\kappa_2|)}.$$

Similarly, to the bipolar case, we can note K_β in the girdle case is also a sort of dissimilarity measure between absolute values of concentration degrees. It is noticeable that $K_\beta(.,.) \in (0, 1]$ such that $K_\beta(\kappa_1, \kappa_2) = 1$ if only if $\kappa_1 = \kappa_2$ and $K_\beta(.,.) \in [1, \infty)$ for $\beta > 1$. This last result is formalized in subsequent corollary. The measure $K_\beta(.,.)$ is illustrated in Figure 8. Moreover, it follows from the definition that $K_{\beta_1}(\kappa_1, \kappa_2) = K_{1-\beta_1}(\kappa_2, \kappa_1)$ for $\beta_1 \in (0, 1/2)$.

Our objective was to propose a hypothesis test via permutation test and bootstrap test being the test statistic used based on Renyi distance denoted by $\boldsymbol{S}_{\boldsymbol{R}:\beta}$ whose formula is given by

$$\boldsymbol{S}_{\boldsymbol{R},\beta} \equiv \frac{2N_1N_2}{N_1+N_2}\frac{1}{\beta(\beta-1)}\log[K_\beta(\hat{\kappa}_1, \hat{\kappa}_2)]. \tag{4.2}$$

4.3 Permutation Tests and Bootstrap

Bootstrap methods and permutation tests based on pivotal statistics were proposed by Amaral et al. (2007) in directional statistics and shape analysis. The bootstrap methodology was proposed by Efron (1979) and was used by Fisher and Hall (1989) and Fisher et al. (1996) for constructing bootstrap confidence regions based on pivotal statistics with directional data. The permutation tests, widely used in multi-sample problems were proposed by Wellner (1979) for directional data.

We considered two Watson populations $W_q(\boldsymbol{\mu}_1, \kappa_1)$ and $W_q(\boldsymbol{\mu}_2, \kappa_2)$, where the concentration parameters κ_1 and κ_2 are known. An extensive simulation study was undertaken and we present the results for the dimension of the sphere to test $H_0 : \pm\mu_1 = \pm\mu_2 = \pm\mu$. This study was carried out for investigating the performance of the test for the statistic given by (4.2), the bootstrap test and the permutation test. First estimated levels of significance obtained for a nominal significance level of 5% of the two tests and second, we determined the empirical power of the tests.

The algorithm for implementing the permutation test can be described in the following four steps:

1. Compute the observed test statistic $\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) = \hat{\theta}(\boldsymbol{Z}, \nu)$.

2. For each replicate, indexed $b = 1, \cdots, B$:
   (a) Generate a random permutation $\pi_b = \pi_\nu$.
   (b) Compute the statistic $\hat{\theta}^{(b)} = \hat{\theta}^*(\boldsymbol{Z}, \pi_b)$.

3. If large values of $\hat{\theta}$ support the alternative, compute the ASL (the empirical p-value) by

$$\hat{p} = \frac{1 + \#\{\hat{\theta}^{(b)} \geq \hat{\theta}\}}{B+1} = \frac{1 + \sum_{b=1}^{B} I(\hat{\theta}^{(b)} \geq \hat{\theta})}{B+1}.$$

   For a lower-tail or two-tail test $\hat{p}$ is computed in a similar way.

4. Reject H_0 at significance level α if $\hat{p} \leq \alpha$.

The algorithm for the bootstrap test can be implemented in the following steps:

1. Compute the observed test statistic $\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) = \hat{\theta}(\boldsymbol{Z}, \nu)$.

2. For each replicate, indexed $b = 1, \cdots, B$:
   (a) Generate a bootstrap sample $\pi_b = \pi_\nu$.
   (b) Compute the statistic $\hat{\theta}^{(b)} = \hat{\theta}^*(\boldsymbol{Z}, \pi_b)$.

3. If large values of $\hat{\theta}$ support the alternative, compute the ASL (the empirical p-value) by

$$\hat{p} = \frac{1 + \#\{\hat{\theta}^{(b)} \geq \hat{\theta}\}}{B+1} = \frac{1 + \sum_{b=1}^{B} I(\hat{\theta}^{(b)} \geq \hat{\theta})}{B+1}.$$

   For a lower-tail or two-tail test $\hat{p}$ is computed in a similar way.

4. Reject H_0 at significance level α if $\hat{p} \leq \alpha$.

4.4 Results

4.4.1 Simulation study

We present a simulation study in order to assess the three proposed hypothesis tests for the Watson distribution concentration parameter κ. The parameter κ is very important in axial data analysis and its interpretation is directly linked to the dispersion degree of these data, such as illustrated in Figure 9. It is noticeable that when the value of κ increases, the data becomes more concentrated in poles $\kappa > 0$ or the equator $\kappa < 0$.

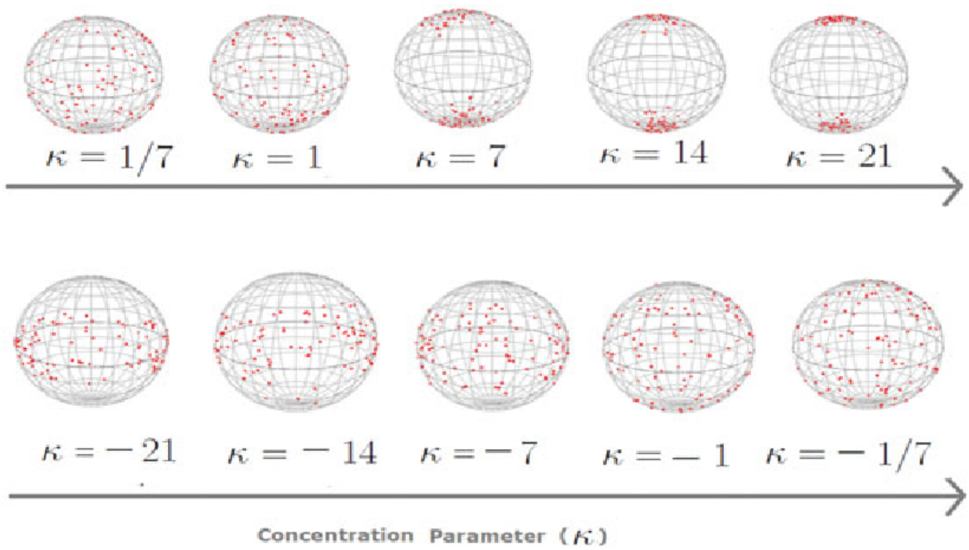


Figure 9 – Illustration of the concentration parameter dynamics of the Watson distribution. (a) Bipolar case and (b) Girdle case.

We considered without loss of generality, that under $H_0 : \pm\mu_1 = \pm\mu_2 = \pm\mu$, where $\mu = (0, 0, 1)^\top$. We generated two samples of sizes n1 and n2 of the populations $W_q(\mu, \kappa_1)$ and $W_q(\mu, \kappa_2)$, assuming the sample sizes $n = 20, 30$ and 50. The estimated levels of significance obtained for a nominal significance level of 5% are indicated in Tables 3 and 4 for known and equal concentration parameters ($\kappa_1 = \kappa_2 = \kappa$) in bipolar case and girdle case, respectively.

In these tables the levels of significance between 4.5% and 5.5%, that may be considered close to the nominal level 5%. Each estimated significance level, i.e., the proportion of times that H_0 is incorrectly rejected, was obtained through a simulation study with 1000 Monte Carlo simulations in the bootstrap and permutation tests. The number of bootstrap re-samples, B, in each Monte Carlo simulation was $B = 500$ and the number of permutation samples was $C = 500$. The significance level for the permutation and bootstrap test is 0.05.

The significance levels obtained in the permutation test in the case of equal concentration parameters are very close to the nominal significance level 5% in almost all cases. Consequently, the permutation test is generally very reliable in what concerns the type I error. From the estimated significance levels obtained in the bootstrap test, we conclude that the bootstrap statistic is very reliable in most part of the considered cases. Additionally, in general, the bootstrap test has similar accuracy to the permutation test, essentially in the case of equal concentration parameters.

Table 3 – Estimated significance levels (in %) of permutation and bootstrap tests, for concentration parameter κ and several sample sizes $n1, n2$.

κ_x	κ_y	n	$S^{\boldsymbol{R}}_{0.3}$		$S^{\boldsymbol{R}}_{0.5}$		$S^{\boldsymbol{R}}_{0.8}$	
			Perm.	Boot.	Perm.	Boot.	Perm.	Boot.
2	2	20	4.6	5.7	3.1	5.3	5.3	5.8
		30	4.4	4.4	5.6	5.1	5.3	5.3
		50	4.6	4.0	4.7	4.7	5.1	4.8
4	4	20	3.4	5.5	5.3	5.2	6.0	5.4
		30	3.7	5.5	5.1	5.3	6.4	5.2
		50	3.9	5.8	5.8	5.8	5.8	5.6
7	7	20	5.7	5.2	4.7	4.6	5.2	5.2
		30	5.0	5.1	4.6	4.8	5.5	5.3
		50	4.8	4.9	4.9	5.2	5.6	5.3
10	10	20	5.6	6.3	2.8	5.5	5.1	5.1
		30	4.3	5.6	5.1	4.0	5.3	5.2
		50	4.8	5.4	5.2	5.3	5.3	5.2

Table 4 – Estimated significance levels (in %) of permutation and bootstrap tests , for concentration parameter κ and several sample sizes $n1, n2$.

κ_x	κ_y	n	$S^{\boldsymbol{R}}_{0.3}$		$S^{\boldsymbol{R}}_{0.5}$		$S^{\boldsymbol{R}}_{0.8}$	
			Perm.	Boot.	Perm.	Boot.	Perm.	Boot.
-2	-2	20	5.3	5.3	3.7	5.8	5.6	4.3
		30	5.3	5.3	3.7	5.2	5.3	4.6
		50	5.1	4.8	4.0	5.2	4.9	3.8
-4	-4	20	3.9	5.8	5.8	5.8	5.8	5.6
		30	3.7	6.4	5.2	5.8	6.3	5.2
		50	4.7	5.1	4.8	4.3	5.3	4.8
-7	-7	20	5.3	5.1	5.7	5.6	5.2	5.8
		30	5.2	5.1	5.3	5.3	5.3	5.3
		50	4.4	5.3	5.2	3.9	5.4	5.2
-10	-10	20	4.0	5.3	5.2	3.2	5.3	5.3
		30	3.7	5.5	5.2	3.5	7.1	5.2
		50	3.6	6.3	5.3	2.9	5.2	5.3

We determined the empirical power of the tabular, bootstrap and permutation tests for a nominal significance level of 5%. We supposed the same null hypothesis as before and in the alternative hypothesis H_1, two directional parameters $\pm\mu_1$ and $\pm\mu_2$ and without loss of generality, we generated one sample from $W_q(\mu, \kappa_1)$ and the other sample from $W_q(\mu, \kappa_2)$.

In the bootstrap or permutation test, the empirical power was obtained from 1000 Monte Carlo simulations, where in each simulation, two samples were generated under H_1 and 500 bootstrap or permutation re-samples were considered.

We indicate the empirical power of the tests for different and known concentration parameters in Table 5 and Table 6. In these tables the values of the power, in which the bootstrap test is more powerful than the permutation tests.

The empirical power of this test not remains close to the significance level for low values of the concentration parameters. For large values of the concentration parameters, the permutation test has better performance for equal-sized samples. we also observed that for samples of equal size, the empirical power of the permutation test increases as the concentration parameters increase.

Table 5 – Empirical power (in %) permutation and bootstrap tests for distinct values of the concentration parameter.

κ_x	κ_y	n	$S_{0.3}^{\boldsymbol{R}}$		$S_{0.5}^{\boldsymbol{R}}$		$S_{0.8}^{\boldsymbol{R}}$	
			Perm.	Boot.	Perm.	Boot.	Perm.	Boot.
2	4	20	15.8	21.9	6.5	35.8	12.9	36.5
		30	18.3	54.0	10.1	56.0	16.4	56.5
		50	22.7	63.7	11.7	66.1	18.6	70
2	6	20	14	19.8	17.2	45.7	18.4	36.2
		30	15.4	52.8	18.3	54.8	17.1	56.7
		50	21.8	80.5	24.2	82.2	28.7	82.9
2	7	20	16.2	26.9	19.6	37.7	18.3	21.1
		30	18.3	28.1	20	44.2	19.2	55.8
		50	21.1	63.3	27.2	65.4	28.1	80.0
2	10	20	17.3	26.2	16.9	27.9	19.7	26.0
		30	19.8	33.6	21.1	46.6	21.1	75.7
		50	22.6	90.2	28.5	86.6	29.2	96.7

Table 6 – Empirical power (in %) permutation and bootstrap tests for distinct values of the concentration parameter.

κ_x	κ_y	n	$S^{\boldsymbol{R}}_{0.3}$		$S^{\boldsymbol{R}}_{0.5}$		$S^{\boldsymbol{R}}_{0.8}$	
			Perm.	Boot.	Perm.	Boot.	Perm.	Boot.
-2	-4	20	18.6	21.6	23.4	42.2	26.5	32.8
		30	12.6	24.1	26.0	46.8	29.1	45.4
		50	22.4	63.5	27	66.1	28.6	72
-2	-6	20	18.4	28.3	16.3	48.2	28.7	32.2
		30	15.4	32.8	18.3	54.8	28.1	56.7
		50	27.5	61.0	27.1	70.8	28.8	73.1
-2	-7	20	18.4	35.5	21.0	46.2	23.3	59.7
		30	23.7	56.7	24.2	55.7	26.2	67.9
		50	21.1	63.8	29.2	65.4	28.1	81.9
-2	-10	20	19.9	63.7	24.8	57.6	25.7	77.1
		30	24.5	70.8	25.8	68.0	26.9	87.1
		50	28.2	90.9	28.4	96.5	28.6	97.0

Figure 10, Figure 11, Figure 12, Figure 13, Figure 14, Figure 15, Figure 16 and Figure 17 shows the empirical power of the tests for different and known concentration parameters.

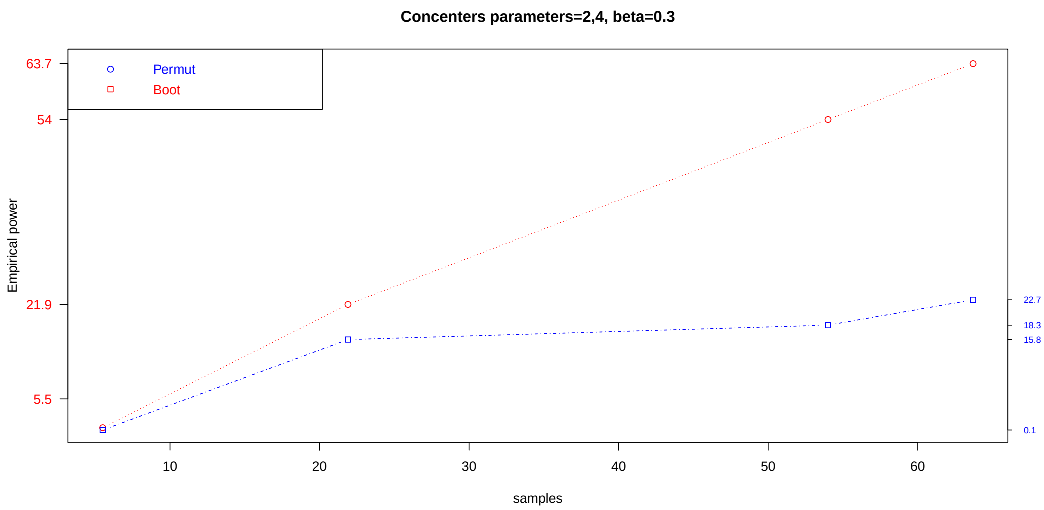


Figure 10 – Empirical power of the tests for estimated different concentrations=2,4 with beta=0.3.

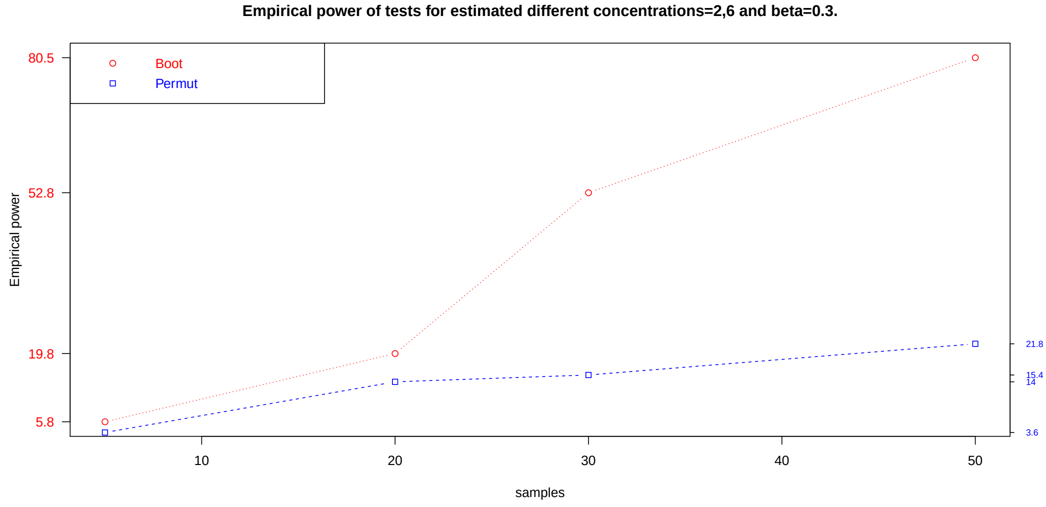


Figure 11 – Empirical power of the tests for estimated different concentrations=2,6 with beta=0.3.

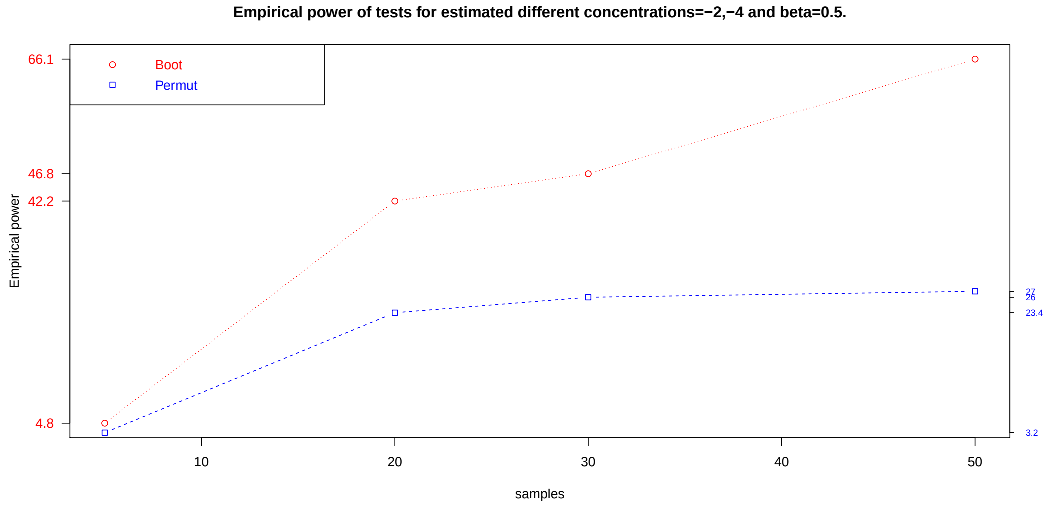


Figure 12 – Empirical power of the tests for estimated different concentrations=-2,-4 with beta=0.5.

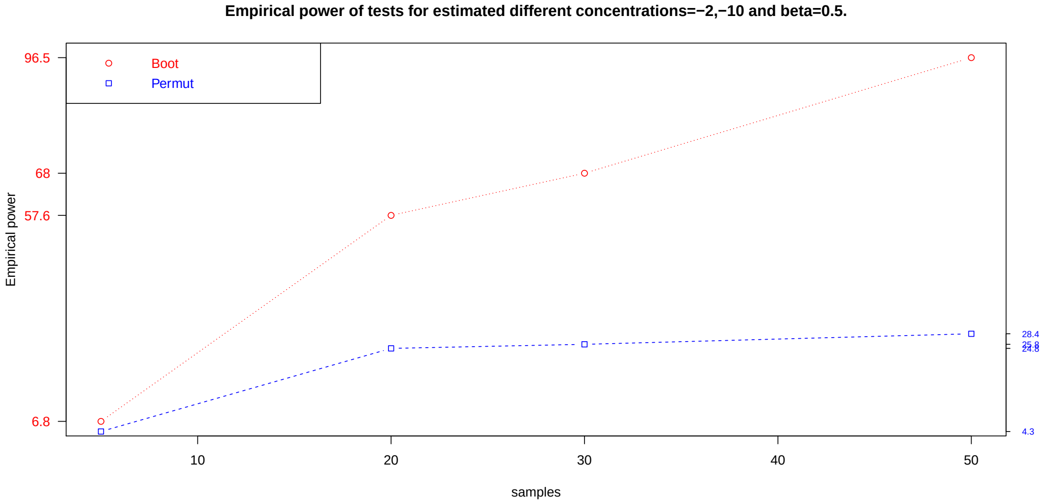


Figure 13 – Empirical power of the tests for estimated different concentrations=-2,-10 with beta=0.5.

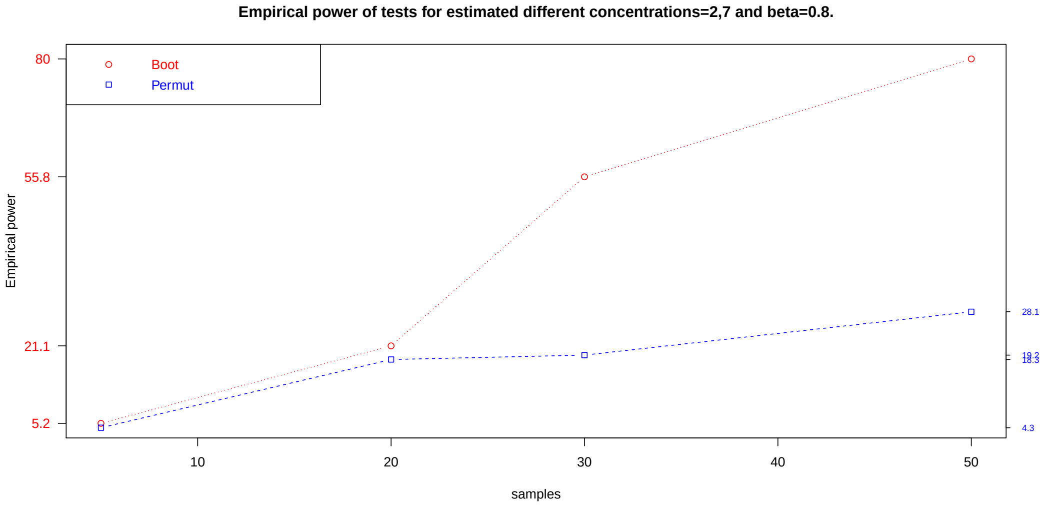


Figure 14 – Empirical power of the tests for estimated different concentrations=2,7 with beta=0.8.

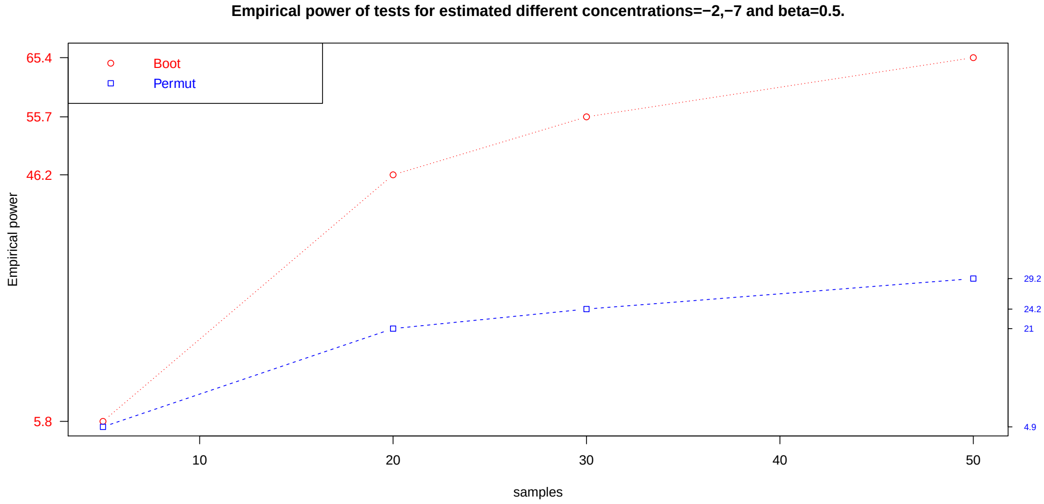


Figure 15 – Empirical power of the tests for estimated different concentrations=-2,-7 with beta=0.5.

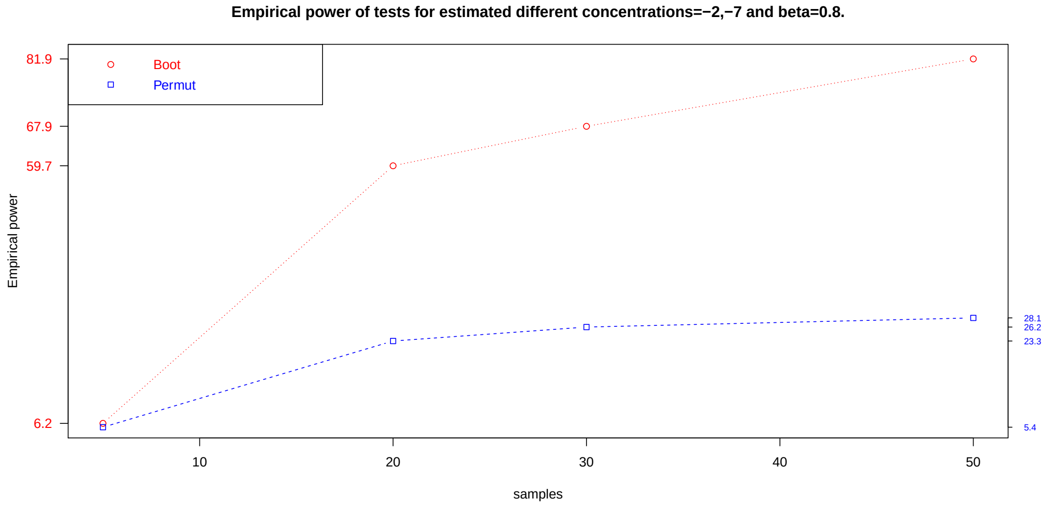


Figure 16 – Empirical power of the tests for estimated different concentrations=-2,-7 with beta=0.8.

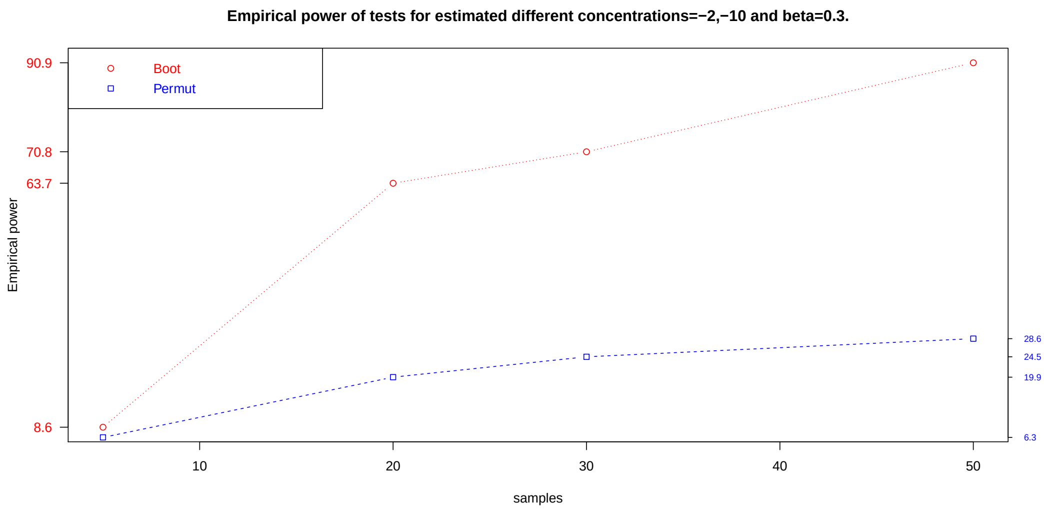


Figure 17 – Empirical power of the tests for estimated different concentrations=-2,-10 with beta=0.3.

4.4.2 Application to real axial data

We perform an application to axes obtained from a sociological study of the attitudes of 48 individuaus to 16 different occupations , judgments were made according 4 different criteria (Earnings, Social Sttatus, Reward, Social Usefulness), given rise to 4 samples (each of 48 multivariate measurements), by Fisher, Lewis, and Embleton (1987). From so-called external analysis of the occupational judgments, each multivariate measurement was reduced to a (spherical) unit vector, yielding the 4 samples of unit vectors.

They quantified simple are represented by latitude θ and longitude ϕ on two populations, based on these angles presented vectors on the unit sphere can be obtained by the transformation: $x = \sin(\theta)\cos(\phi)$, $y = \sin(\theta)\sin(\phi)$ and $z = \cos(\theta)$. For the study of real data analysis we use only 2 criteria, in this case earnings and social status both criteria can be seen in figures 18 and 19.

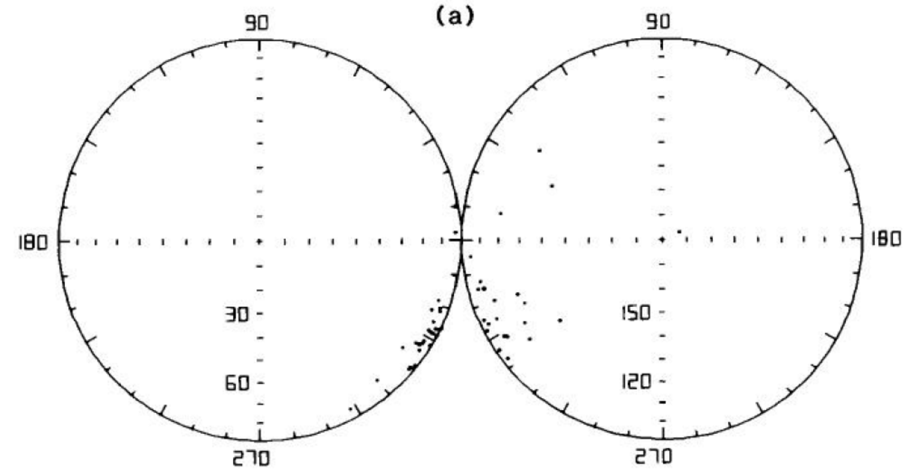


Figure 18 – Equal-area projections of occupational judgments according earnings criteria.

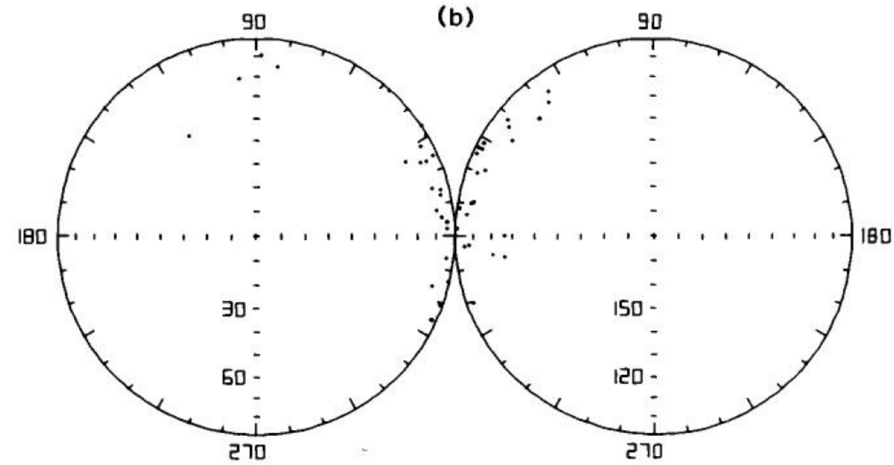


Figure 19 – Equal-area projections of occupational judgments according social standing criteria

Table 7 – Empirical power (in %) of permutation tests and bootstrap, for different concentration and several sample sizes $n1, n2$.

Criteria	Concentration	Statistic	p-value(%)	
			Perm.	Boot.
Earnings	Equal	1.297	20.7	51.3
Social Status	Equal	2.673	10.8	39.5

Table 7 as well as the p-values obtained for the tabular method, the bootstrap and permutation versions of Renyi distance statistic. The p-values of the bootstrap and permutation tests were obtained with B = 1000 bootstrap re-samples and C = 1000 permutation samples. First, the difference between the p-values of the tests for both statistics is very small. Second, on one hand, the three tests led to the same conclusion for tow criteria. More precisely, we can conclude that two is no significant difference between the systems for earnings criteria. On the other hand, for social status there is no evidence to conclude that the systems differ using the Renyi distance statistic and bootstrap tests. Based on the permutation test, we conclude that there is difference between the systems at a level of 5%.

4.5 Conclusion

We have concluded that the bootstrap and permutation versions of the Renyi distance statistic for testing a common mean polar axis across several Watson populations defined on the hypersphere gave reliable estimates of the significance level, in most part of the simulated cases, and in particular, for small concentrations and small samples. Additionally, from the two tests, the bootstrap test is in general the most powerful test in the case of small samples for small concentrations between the Watson populations. So, in these cases, the bootstrap and permutation tests based on Renyi distance statistic may constitute useful alternatives to this statistic, that has an asymptotic distribution, valid only for large concentrations.

5 CONCLUSION

We summarize the main contributions of this thesis in the following items:

- The chapter 2 has presented a bias correction estimation method for the concentration parameter at the W model. It was made according to the Cox-Snell methodology. We have proposed closed-form expressions the Fisher information matrix and the second-order bias of ML estimator of concentration parameter. We concluded that in both cases the performance of our estimate for bias correction showed satisfactory performance and a significant reduction of MSE for the different sample sizes.

- In Chapter 3 we proposed the derivation of the quadratic statistical distance for directional data; in particular, following the real Watson distribution and various theoretical properties for this model.
- In Chapter 4, we performed hypothesis testing using permutation and bootstrap testing for two real Watson distribution samples, based of the Rényi distance statistic. We conclude that the bootstrap and the permutation test showed estimates very close to the confidence level considered, but as for the power of the test in all cases considered the bootstrap presents better performance compared to the permutation test.

Several lines of research can still be addressed, such as:

- To present a bias correction estimation method for the concentration parameter in the W model, using the Bartlett methodology.

- Propose a new hypothesis test to verify if two samples from Watson populations distributed with different dominant axes are equally concentrated, using divergence function class.

In closing this work, we hope to have made a relevant contribution to the research that addresses Bias-corrected estimation for the real Watson model, the derivation of the quadratic statistical distance for directional data with theoretical properties and Bootstrap and Permutation Tests for two sample problems of axial data analysis.

REFERENCES

[1] Abramowitz, M., and I. A. Stegun. 1994. *Handbook of mathematical functions.* New York: Dover Publications.

[2] Agostinelli, C. 2007. Robust estimation for circular data. *Comput. Stat Data Anal.* 51:5867–5875.

[3] Akaike, H. 1954. An approximation to the density function. *Annals of the Institute of Statistical Mathematics* 6:127-132.

[4] Amaral, G. J. A., Dryden, I. L. and Wood, A. T. A. 2007. Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. *Journal of the American Statistical Association* 102: 695-707.

[5] Anderson, C. M. and Wu, C. F. J. 1995. Measuring location effects from factorial experiments with a directional response. *International Statistical Review* 63: 345-363.

[6] Bai, Z., Rao, C., Zhao, L. 1988. Kernel estimators of density of directional data. *Journal of Multivariate Analysis* 27:24–39.

[7] Barros, C. M., G. J. A. Amaral, A. D. C. Nascimento, and A. H. M. A. Cysneiros. 2016. Detecting influential observations in Watson data. *Communications in Statistics — Theory and Methods* 0:1–17.

[8] Basu, A., Lindsay, B. 1994. Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Ann Inst Stat Math* 46:683–705.

[9] Beran, R. 1977. Exponential models for directional data. *The Annals of Statistics* 7:1162–1178.

[10] Best, D. J., and N. I. Fisher. 1986. Goodness-of-fit and discordancy tests for samples from the Watson distribution on a sphere. *Australian Journal of Statistics* 28:13–31.

[11] Best, D. J., and N. I. Fisher. 1981. The bias of the maximum likelihood estimators of the von mises-fisher concentration parameters.. *Communications in Statistics - Simulation and Computation* 5:493–502.

[12] Bingham, C. 1974. An antipodally symmetric distribution on the sphere. *The Annals of Statistics* 2:1201–5.

[13] Cao, R., Cuevas, A., Manteiga, W. G. 1994. A comparative study of several smoothing methods in density estimation. *Computation Statistics Data Analysis* 17:153–176.

[14] Cao, R., Cuevas, A., Fraiman, R. 1995. Minimum distance density-based estimators. *Comput. Stat Data Anal.* 20:611–631.

[15] Cordeiro, G. M., and F. Cribari. 2014. *An Introduction to Bartlett Correction and Bias Reduction*. New York: Springer Heidelberg.

[16] Cox, D. R., and E. J. Snell. 1968. A General Definition of Residuals. *Journal of the Royal Statistical Society* 30:248–275.

[17] Cox, D. 2000. *Theoretical Statistics*. London: Chapman & Hall.

[18] Dhillon, I. S. Marcotte, E. M. and Roshan, U. 2003. Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics* 52:1612–1619.

[19] Dore, L. H. G., G. J. A. Amaral, J .T. M. Cruz, and A.T. A. Wood. 2016. Bias-corrected maximum likelihood estimation parameters of the complex Bingham distribution. *Brazilian Journal of Probability and Statistics* 30:385–400.

[20] Efron, B. (1979). Bootstrap methods: another look at the jacknife. *The Annals of Statistics* 7: 1-26.

[21] Embleton, B., Giddings, J. 1974. Late Precambrian and lower Paleozoic palaeomagnetic results from South Australia and Western Australia. *Earth Planet Sci Lett* 22:355–365.

[22] Ferguson, D. 1967. Sun compass orientation of the northern cricket frog Acris crepitans. *Anim Behav* 14:45–53.

[23] Figueiredo, A. 2007. Discordancy tests based on the likelihood ratio for the bipolar Watson distribution on the hypersphere. *Communications in Statistics: Simulation and Computation* 36: 413–21.

[24] Fisher, N. I. Lewis, T. and Embleton, B. J. J. 1987. *Statistical analysis of spherical data*. Cambridge: Cambridge University Press.

[25] Fisher, N. I., B. J. J. Embleton, and T. Lewis. 1993. *Statistical Analysis of Spherical Data*. Cambridge: Cambridge University Press.

[26] Frery, A. C., A. D. C. Nascimento, and R. J. Cintra. 2014. Analytic expressions for stochastic distances between relaxed complex Wishart distributions. *IEEE Transactions on Geoscience and Remote Sensing* 52:1213–26.

[27] García, P. E. 2013. Exact risk improvement of bandwith selectors for kernel density estimation with directional data. *Electron Journal Statistics* 7:1655-1685.

[28] Hall, P., Watson, G., Cabrera, J. 1987. Kernel density estimation with spherical data. *Biometrika* 74:751–762.

[29] He, X. 1992. Robust statistics of directional data: a survey. *Nonparametric Statistics and Related Topics* 17:87–95.

[30] Hering, A., Genton, M. 1987. Powering up with space-time wind forecasting. *Journal of the American Statistical Association* 105:92–104.

[31] Jammalamadaka, S. R. and SenGupta, A. 2001. *Topics in Circular Statistics.* World Scientific: Singapore.

[32] Kent, J. T. 1994. The complex Bingham distribution and shape and analysis. *Journal of the Royal Statistical Society: Series B (Methodological)* 56:285–99.

[33] Kim-Hung, L., and K .W. Carl. 1993. Random sampling from the Watson Distribution. *Random sampling from the Watson Distribution.* 22:997–1009.

[34] Ko, D., Chang, T. 1993. Robust M-estimators on spheres. *Journal Multivariate Analysis* 45:104–136.

[35] Kume, A., and S. G. Walker. 2014. On the Bingham distribution with large dimension. Journal of Multivariate Analysis. *Journal of Multivariate Analysis* 124:345–352.

[36] Kume, A., and A. T. A. Wood. 2005. Saddlepoint approximations for the Bingham and Fisher- Bingham normalising constants. *Biometrika* 92:465–76.

[37] Kurz, G, Gilitschenski, I., Hanebeck, U. D. 2016. Unscented vom Mises-Fisher Filtering *.IIEE Signal Processing Letters* 23:463–467.

[38] Leong, P., Carile, S. 1998. Methods for spherical data analysis and visualization. *Journal Neurosci Methods* 80:191–200.

[39] Li, K. H. and Wong, C. K. F. 1993. Random sampling from theWatson distribution. *Communications in Statistics - Computation and Simulation* 22: 997-1009.

[40] Lindsay, B. G. 1994. Efficiency Versus Robustness: The Case for Minimum Hellinger Distance and Related Methods. *The Annals of Statistics* 22:1081-1114.

[41] Mardia, K., and P. Jupp. 2000. *Directional Statistics.* USA: John Wiley & Sons.

[42] Muller, H. G. 2006. Density estimation-II. *Encyclopedia of statistical sciences* 2:1611–1626.

[43] Nascimento, A. D. C. Silva, R. C. and Amaral, G. J.A. 2018. Distance-based hypothesis tests on the Watson distribution *Communications in Statistics - Simulation and Computation* 1:1-14.

[44] Parzen, E. 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33:1065–1076.

[45] Rosemblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27:832–837.

[46] Salicrú, M., M. L. Menéndez, L. Pardo, and D. Morales. 1994. On the applications of divergence type measures in testing statistical hypothesis. *Journal of Multivariate Analysis* 51:372–91.

[47] Silverman, B. W. 1986. *Density estimation of statistics and data analysis.* London: Chapman & Hall.

[48] Simpson, D. G. 1989. Hellinger Deviance Tests: Efficiency, breakdown points, and examples. *Journal of the American Statistical Association* 84:107-113.

[49] Small, C. G. 1996. *The Statistical Theory of Shape.* Germany: Springer-Verlag.

[50] Sra, S., and D. Karp. 2013. The multivariate Watson distribution: Maximum-likelihood estimation and other aspects. *Journal of Multivariate Analysis* 114:256–269.

[51] Stephens, M. A. (1969). Multi-sample tests for the Fisher distribution for directions. *Biometrika* 56:169-181.

[52] Taylor, C. 1994. Automatic bandwidth selection for circular density estimation. *Comput. Stat Data Anal.* 52:3493–3500.

[53] Underwood, A. J. and Chapman, M. G. (1985). Multifactorial analysis of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology* 91:17-43.

[54] Watson, G. S. 1965. Equatorial distributions on a sphere. *Biometrika* 52:193–201.

[55] Wellner, J. A. (1979). Permutation tests for directional data. *The Annals of Statistics* 7 929-943.

[56] Zhao, L., Wu, C. 2001. Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser.* 44:474–473.

APPENDIX A – COMPONENTS OF THE FISHER INFORMATION MATRIX

Consider to derive the expression (2.4). According to Mardia and Jupp (2000) if $\mathbf{\Sigma}$ is the parameter of the real Bingham model, the W distribution may be obtained the former from the identity

$$\mathbf{\Sigma}_p = (\boldsymbol{I}_p - 2\kappa\boldsymbol{\mu}\boldsymbol{\mu}^\top)^{-1}.$$

Thus

$$\mathbf{\Sigma}_3=(\mathbb{I}_3-2\kappa\boldsymbol{\mu}\boldsymbol{\mu}^\top)^{-1}=\frac{1}{1-2\kappa\mu_3^2-2\kappa\mu_2^2-2\kappa\mu_1^2}\begin{pmatrix} 1-2\kappa\mu_3^2-2\kappa\mu_2^2 & 2\kappa\mu_1\mu_2 & 2\kappa\mu_1\mu_3 \\ 2\kappa\mu_1\mu_2 & 1-2\kappa\mu_3^2-2\kappa\mu_1^2 & 2\kappa\mu_2\mu_3 \\ 2\kappa\mu_1\mu_3 & 2\kappa\mu_2\mu_3 & 1-2\kappa\mu_2^2-2\kappa\mu_1^2 \end{pmatrix} \quad \text{(A.1)}$$

and, as consequence,

$$\operatorname{tr}(\mathbf{\Sigma_3}) = \frac{3-4\kappa\mu_3^2-4\kappa\mu_2^2-4\kappa\mu_1^2}{1-2\kappa\mu_3^2-2\kappa\mu_2^2-2\kappa\mu_1^2}. \quad \text{(A.2)}$$

Replacing the equations (8) and (9) into (6), we have

$$\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) \approx \frac{1}{p+1} + \frac{2\mathbf{\Sigma}}{(p+1)(p+3)} + \frac{2\,\mathbb{I}_p\operatorname{tr}(\mathbf{\Sigma})}{(p+1)^2(p+3)} \approx$$

$$\approx \begin{pmatrix} \frac{13-28\kappa\mu_3^2-28\kappa\mu_2^2-20\kappa\mu_1^2}{48(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{\kappa\mu_1\mu_2}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{\kappa\mu_1\mu_3}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} \\ \frac{\kappa\mu_1\mu_2}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{13-28\kappa\mu_3^2-20\kappa\mu_2^2-28\kappa\mu_1^2}{48(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{\kappa\mu_2\mu_3}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} \\ \frac{\kappa\mu_1\mu_3}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{\kappa\mu_2\mu_3}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} & \frac{13-20\kappa\mu_3^2-28\kappa\mu_2^2-28\kappa\mu_1^2}{48(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} \end{pmatrix}.$$

Finally, after some algebraic manipulations the FIM components are given by

- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_1\partial\mu_1}\right) = \mathbb{E}(2\kappa\boldsymbol{S}_{11}) \approx \frac{13\kappa-28\kappa^2\mu_3^2-28\kappa^2\mu_2^2-20\kappa^2\mu_1^2}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_1\partial\mu_2}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_2\partial\mu_1}\right) = \mathbb{E}[\kappa(\boldsymbol{S}_{12}+\boldsymbol{S}_{21})] \approx \frac{\kappa^2\mu_1\mu_2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_1\partial\mu_3}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_3\partial\mu_1}\right) = \mathbb{E}[\kappa(\boldsymbol{S}_{13}+\boldsymbol{S}_{31})] \approx \frac{\kappa^2\mu_1\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_1\partial\kappa}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\kappa\partial\mu_1}\right) = \mathbb{E}(2\boldsymbol{S}_{11}\mu_1+\boldsymbol{S}_{21}\mu_2+\boldsymbol{S}_{31}\mu_3+\boldsymbol{S}_{12}\mu_2+\boldsymbol{S}_{13}\mu_3) \approx \frac{13\mu_1-20\kappa\mu_1\mu_3^2-20\kappa\mu_1\mu_2^2-20\kappa\mu_1^3}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_2\partial\mu_2}\right) = \mathbb{E}(2\kappa\boldsymbol{S}_{22}) \approx \frac{13\kappa-28\kappa^2\mu_3^2-20\kappa^2\mu_2^2-28\kappa^2\mu_1^2}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_2\partial\mu_3}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_3\partial\mu_2}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_3\partial\mu_2}\right) = \mathbb{E}[\kappa(\boldsymbol{S}_{32}+\boldsymbol{S}_{23})] \approx \frac{\kappa^2\mu_2\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,

- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_2\partial\kappa}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\kappa\partial\mu_2}\right) = \mathbb{E}(\boldsymbol{S}_{21}\mu_1+\boldsymbol{S}_{12}\mu_1+2\boldsymbol{S}_{22}\mu_2+\boldsymbol{S}_{32}\mu_3+\boldsymbol{S}_{23}\mu_3) \approx \frac{13\mu_2-20\kappa\mu_1^2\mu_2-20\kappa\mu_2\mu_3^2-20\kappa\mu_2^3}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_3\partial\mu_3}\right) = \mathbb{E}(2\kappa\boldsymbol{S}_{33}) \approx \frac{13\kappa-20\kappa^2\mu_3^2-28\kappa^2\mu_2^2-28\kappa^2\mu_1^2}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$,
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\mu_3\partial\kappa}\right) = \mathbb{E}\left(\frac{-\partial^2\ell}{\partial\kappa\partial\mu_3}\right) = \mathbb{E}(\boldsymbol{S}_{31}\mu_1+\boldsymbol{S}_{32}\mu_2+\boldsymbol{S}_{13}\mu_1+\boldsymbol{S}_{23}\mu_2+2\kappa\boldsymbol{S}_{33}\mu_3) \approx \frac{13\mu_3-20\kappa\mu_3^3-20\kappa\mu_1^2\mu_3-20\kappa\mu_2^2\mu_3}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)}$ and
- $\mathbb{E}\left(\frac{-\partial^2\ell}{\partial\kappa\partial\kappa}\right) = \frac{0.5^2(2.5)M(1.5,2.5,\kappa)^2-0.5(1.5)^2M(0.5,1.5,\kappa)M(2.5,3.5,\kappa)}{1.5^2(2.5)M(0.5,1.5,\kappa)^2}$.

APPENDIX B – TERMS USED IN THE ESTIMATION OF THE BIAS CORRECTION OF PARAMETER κ

- $\kappa_{11}^{(1)} = \frac{3\kappa^2\mu_1-8\kappa^3\mu_1{\mu_3}^2-8\kappa^3\mu_1{\mu_2}^2}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_1}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_1}{\mathbf{D}_1}$
- $\kappa_{11}^{(2)} = \frac{8\kappa^3\mu_1^2\mu_2-\kappa^2\mu_2}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_2}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_2}{\mathbf{D}_1}$
- $\kappa_{11}^{(3)} = \frac{8\kappa^3\mu_1^2\mu_3-\kappa^2\mu_3}{6(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_3}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_3}{\mathbf{D}_1}$
- $\kappa_{11}^{(4)} = \frac{13-56\kappa{\mu_3}^2-56\kappa{\mu_2}^2-40\kappa{\mu_1}^2+56\kappa^2{\mu_3}^4+112\kappa^2{\mu_2}^2{\mu_3}^2+96\kappa^2{\mu_1}^2{\mu_2}^2+96\kappa^2{\mu_1}^2{\mu_3}^2+56\kappa^2{\mu_2}^4+40\kappa^2{\mu_1}^4}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_4}{\mathbf{D}_1}$
- $\kappa_{114} = \frac{13-28\kappa\mu_3^2-28\kappa\mu_2^2-20\kappa\mu_1^2}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} = \frac{\mathbf{N}_5}{\mathbf{D}_2} = \kappa_{141} = \kappa_{411}$
- $\kappa_{12}^{(1)} = \frac{\kappa^2\mu_2-2\kappa^3\mu_2{\mu_3}^2-2\kappa^3{\mu_2}^3+2\kappa^3\mu_1^2\mu_2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_6}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_6}{\mathbf{D}_1} = \kappa_{21}^{(1)}$
- $\kappa_{12}^{(2)} = \frac{\kappa^2\mu_1-2\kappa^3\mu_1{\mu_3}^2-2\kappa^3{\mu_1}^3+2\kappa^3\mu_1\mu_2^2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_7}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_7}{\mathbf{D}_1} = \kappa_{21}^{(2)}$
- $\kappa_{12}^{(3)} = \frac{4\kappa^3\mu_1\mu_2\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_8}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_8}{\mathbf{D}_1} = \kappa_{13}^{(2)} = \kappa_{21}^{(3)} = \kappa_{23}^{(1)} = \kappa_{31}^{(2)} = \kappa_{32}^{(1)}$
- $\kappa_{12}^{(4)} = \frac{2\kappa\mu_1\mu_2-2\kappa^2\mu_1\mu_2{\mu_3}^2-2\kappa^2\mu_1{\mu_2}^3-2\kappa^2\mu_1 3\mu_2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_9}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_9}{\mathbf{D}_1} = \kappa_{21}^{(4)}$
- $\kappa_{124} = \frac{\kappa\mu_1\mu_2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} = \frac{\mathbf{N}_{10}}{\mathbf{D}_2/8} = \frac{8\mathbf{N}_{10}}{\mathbf{D}_2} = \kappa_{142} = \kappa_{214} = \kappa_{241} = \kappa_{412} = \kappa_{421}$
- $\kappa_{13}^{(1)} = \frac{\kappa^2\mu_3-2\kappa^3{\mu_3}^3-2\kappa^3{\mu_2}^2\mu_3+2\kappa^3\mu_1^2\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{11}}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_{11}}{\mathbf{D}_1} = \kappa_{31}^{(1)}$
- $\kappa_{13}^{(3)} = \frac{\kappa^2\mu_1-2\kappa^3\mu_1{\mu_2}^2-2\kappa^3{\mu_1}^3+2\kappa^3\mu_1\mu_3^2}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\overline{\mathbf{N}}_{12}}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_{12}}{\mathbf{D}_1} = \kappa_{31}^{(3)}$
- $\kappa_{13}^{(4)} = \frac{2\kappa\mu_1\mu_3-2\kappa^2\mu_1{\mu_3}^3-2\kappa^2\mu_1{\mu_2}^2\mu_3-2\kappa^2\mu_1^3\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{13}}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_{13}}{\mathbf{D}_1} = \kappa_{31}^{(4)}$
- $\kappa_{134} = \frac{\kappa\mu_1\mu_3}{3(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)} = \frac{\mathbf{N}_{14}}{\mathbf{D}_2/8} = \frac{8\mathbf{N}_{14}}{\mathbf{D}_2} = \kappa_{143} = \kappa_{314} = \kappa_{341} = \kappa_{413} = \kappa_{431}$
- $\kappa_{14}^{(1)} = \frac{13-46\kappa{\mu_3}^2-46\kappa{\mu_2}^2-34\kappa{\mu_1}^2+40\kappa^2{\mu_3}^4+80\kappa^2{\mu_2}^2{\mu_3}^2+80\kappa^2{\mu_1}^2{\mu_3}^2+80\kappa^2{\mu_1}^2{\mu_2}^2+40\kappa^2{\mu_2}^4+40\kappa^2{\mu_1}^4}{24(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{15}}{\mathbf{D}_1} = \kappa_{41}^{(1)}$
- $\kappa_{14}^{(2)} = \frac{\kappa\mu_1\mu_2}{2(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{16}}{\mathbf{D}_1/12} = \frac{12\mathbf{N}_{16}}{\mathbf{D}_1} = \kappa_{41}^{(2)} = \kappa_{42}^{(1)} = \kappa_{24}^{(1)}$
- $\kappa_{14}^{(3)} = \frac{\kappa\mu_1\mu_3}{2(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{17}}{\mathbf{D}_1/12} = \frac{12\mathbf{N}_{17}}{\mathbf{D}_1} = \kappa_{41}^{(3)} = \kappa_{34}^{(1)} = \kappa_{43}^{(1)}$
- $\kappa_{14}^{(4)} = \frac{\mu_1\mu_3^2+\mu_1\mu_2^2+\mu_3^3}{4(1-2\kappa{\mu_3}^2-2\kappa{\mu_2}^2-2\kappa{\mu_1}^2)^2} = \frac{\mathbf{N}_{18}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{18}}{\mathbf{D}_1}$

- $\kappa_{22}^{(1)} = \frac{8\kappa^3\mu_1\mu_2^2-\kappa^2\mu_1}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{19}}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_{19}}{\mathbf{D}_1}$

  $\kappa_{22}^{(2)} = \frac{3\kappa^2\mu_2-8\kappa^3\mu_2\mu_3{}^2-8\kappa^3\mu_1^2\mu_2}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{20}}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_{20}}{\mathbf{D}_1}$

- $\kappa_{22}^{(3)} = \frac{8\kappa^3\mu_2^2\mu_3-\kappa^2\mu_3}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{21}}{\mathbf{D}/4} = \frac{4\mathbf{N}_{21}}{\mathbf{D}_1}$

- $\kappa_{22}^{(4)} = \frac{13-56\kappa\mu_3{}^2-40\kappa\mu_2{}^2-56\kappa\mu_1{}^2+96\kappa^2\mu_2{}^2\mu_3^2+112\kappa^2\mu_1{}^2\mu_3{}^2+96\kappa^2\mu_1{}^2\mu_2{}^2+56\kappa^2\mu_3{}^4+56\kappa^2\mu_1{}^4+40\kappa^2\mu_2{}^4}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{22}}{\mathbf{D}_1}$

- $\kappa_{224} = \frac{13-28\kappa\mu_3^2-20\kappa\mu_2^2-28\kappa\mu_1^2}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)} = \frac{\mathbf{N}_{23}}{\mathbf{D}_2} = \kappa_{242} = \kappa_{422}$

- $\kappa_{23}^{(2)} = \frac{\kappa^2\mu_3-2\kappa^3\mu_3{}^3-2\kappa^3\mu_1{}^2\mu_3+2\kappa^3\mu_2^2\mu_3}{3(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{24}}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_{24}}{\mathbf{D}_1} = \kappa_{32}^{(2)}$

- $\kappa_{23}^{(3)} = \frac{\kappa^2\mu_2-2\kappa^3\mu_2{}^3-2\kappa^3\mu_1{}^2\mu_2+2\kappa^3\mu_2\mu_3^2}{3(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{25}}{\mathbf{D}_1/8} = \frac{8\mathbf{N}_{25}}{\mathbf{D}_1} = \kappa_{32}^{(3)}$

- $\kappa_{23}^{(4)} = \frac{2\kappa\mu_2\mu_3-2\kappa^2\mu_2\mu_3^3-2\kappa^2\mu_2^3\mu_3-2\kappa^2\mu_1^2\mu_2\mu_3}{3(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{26}}{\mathbf{D}_1/8} = \frac{\mathbf{N}_{26}}{\mathbf{D}_1} = \kappa_{32}^{(4)}$

- $\kappa_{234} = \frac{\kappa\mu_2\mu_3}{3(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)} = \frac{\mathbf{N}_{27}}{\mathbf{D}_2/8} = \frac{8\mathbf{N}_{27}}{\mathbf{D}_2} = \kappa_{243} = \kappa_{324} = \kappa_{342} = \kappa_{423} = \kappa_{432}$

- $\kappa_{24}^{(2)} = \frac{13-46\kappa\mu_1{}^2-46\kappa\mu_3{}^2-34\kappa\mu_2{}^2+80\kappa^2\mu_1{}^2\mu_3^2+40\kappa^2\mu_3^4+80\kappa^2\mu_2{}^2\mu_3{}^2+80\kappa^2\mu_1{}^2\mu_2^2+40\kappa^2\mu_1{}^4+40\kappa^2\mu_2{}^4}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{28}}{\mathbf{D}_1} = \kappa_{42}^{(2)}$

- $\kappa_{24}^{(3)} = \frac{\kappa\mu_2\mu_3}{2(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{29}}{\mathbf{D}_1/12} = \frac{12\mathbf{N}_{29}}{\mathbf{D}_1} = \kappa_{42}^{(3)} = \kappa_{34}^{(2)} = \kappa_{43}^{(2)}$

- $\kappa_{24}^{(4)} = \frac{\mu_1^2\mu_2+\mu_2\mu_3^2+\mu_2^3}{4(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{30}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{30}}{\mathbf{D}_1}$

- $\kappa_{33}^{(1)} = \frac{8\kappa^3\mu_1\mu_3^2-\kappa^2\mu_1}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{31}}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_{31}}{\mathbf{D}_1}$

- $\kappa_{33}^{(2)} = \frac{8\kappa^3\mu_2\mu_3^2-\kappa^2\mu_2}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{32}}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_{32}}{\mathbf{D}_1}$

- $\kappa_{33}^{(3)} = \frac{3\kappa^2\mu_3-8\kappa^3\mu_2^2\mu_3-8\kappa^2\mu_1{}^2\mu_3}{6(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{33}}{\mathbf{D}_1/4} = \frac{4\mathbf{N}_{33}}{\mathbf{D}_1}$

- $\kappa_{33}^{(4)} = \frac{13-40\kappa\mu_3{}^2-56\kappa\mu_2{}^2-56\kappa\mu_1{}^2+96\kappa^2\mu_2{}^2\mu_3^2+40\kappa^2\mu_3^4+96\kappa^2\mu_1{}^2\mu_3{}^2+112\kappa^2\mu_1{}^2\mu_2^2+56\kappa^2\mu_1{}^4+56\kappa^2\mu_2{}^4}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{34}}{\mathbf{D}_1}$

- $\kappa_{334} = \frac{13-20\kappa\mu_3^2-28\kappa\mu_2^2-28\kappa\mu_1^2}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)} = \frac{\mathbf{N}_{35}}{\mathbf{D}_2} = \kappa_{343} = \kappa_{433}$

- $\kappa_{34}^{(1)} = \frac{\kappa\mu_1\mu_3}{2(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{36}}{\mathbf{D}_1/12} = \frac{12\mathbf{N}_{36}}{\mathbf{D}_1} = \kappa_{43}^{(1)}$

- $\kappa_{34}^{(3)} = \frac{13-34\kappa\mu_3{}^2-46\kappa\mu_2{}^2-46\kappa\mu_1{}^2+80\kappa^2\mu_2{}^2\mu_3^2+40\kappa^2\mu_3^4+80\kappa^2\mu_1{}^2\mu_3{}^2+80\kappa^2\mu_1{}^2\mu_2^2+40\kappa^2\mu_1{}^4+40\kappa^2\mu_2{}^4}{24(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{37}}{\mathbf{D}_1} = \kappa_{43}^{(3)}$

- $\kappa_{34}^{(4)} = \frac{\mu_3^2+\mu_2^2+\mu_1^3}{4(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{38}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{38}}{\mathbf{D}_1}$

- $\kappa_{41}^{(4)} = \frac{\mu_1\mu_3^2+\mu_1\mu_2^2+\mu_1^3}{4(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{39}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{39}}{\mathbf{D}_1}$

- $\kappa_{42}^{(4)} = \frac{\mu_2\mu_3^2+\mu_2^3+\mu_1^2\mu_2}{4(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{40}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{40}}{\mathbf{D}_1}$

- $\kappa_{43}^{(4)} = \frac{\mu_3^3+\mu_1^2\mu_3+\mu_2^2\mu_3}{4(1-2\kappa\mu_3{}^2-2\kappa\mu_2{}^2-2\kappa\mu_1{}^2)^2} = \frac{\mathbf{N}_{41}}{\mathbf{D}_1/6} = \frac{6\mathbf{N}_{41}}{\mathbf{D}_1}$
- $\kappa_{44}^{(4)} = \frac{2a^2(a+1)(M(a,b,k))(M(a+1,b+1,k))(M(a+2,b+2,k))-2a^2(a+1)(M(a+2,b+2,k))(M(a+1,b+1,k))}{b^2(b+1)(M(a,b,k))^3} = \frac{\mathbf{N}_{42}}{\mathbf{D}_3}$
- $\kappa_{444} = \frac{2a^2(a+1)(M(a,b,k))(M(a+1,b+1,k))(M(a+2,b+2,k))-2a^2(b+1)(M(a+2,b+2,k))(M(a+1,b+1,k))}{b^2(b+1)(M(a,b,k))^3} = \frac{\mathbf{N}_{43}}{\mathbf{D}_3}.$

APPENDIX C – TERMS USED TO CALCULATE $D_2(\widehat{f}_{h_n}, f_W)$

- $c(h_n) = \frac{5h_n^4}{10\pi h_n^4 - 10\pi - 2\pi a_1^2 - 2\pi a_2^2 - 2a_3^2}$,

- $\int_{\mathbb{S}^2} dx_i = \frac{4\pi}{3}$,

- $\int_{\mathbb{S}^2} \boldsymbol{x}\boldsymbol{x}^\top dx_i = \int_0^1 \int_0^{2\pi} \int_0^{\pi} [a_1 r \sin\theta \cos\theta + a_2 r \sin\theta \sin\phi + a_3 r \cos\theta] r^2 \sin\theta d\theta d\phi dr = 0$,

- $\int_{\mathbb{S}^2} \boldsymbol{x}_i \boldsymbol{x}_i^\top dx_i = \int_0^1 \int_0^{2\pi} \int_0^{\pi} \begin{bmatrix} r\sin\theta\cos\phi \\ r\sin\theta\sin\phi \\ r\cos\theta \end{bmatrix} \begin{bmatrix} r\sin\theta\cos\phi & r\sin\theta\sin\phi & r\cos\theta \end{bmatrix} r^2 \sin\theta d\theta d\phi dr$
  $= \int_0^1 \int_0^{2\pi} \int_0^{\pi} \begin{bmatrix} r^2\sin^2\theta\cos^2\phi & r^2\sin^2\theta\cos\phi\sin\phi & r^2\sin\theta\cos\phi\cos\theta \\ r^2\sin^2\theta\sin\phi\cos\phi & r^2\sin^2\theta\sin^2\phi & r^2\sin\theta\sin\phi\cos\theta \\ r^2\cos\theta\sin\theta\cos\phi & r^2\cos\theta\sin\theta\sin\phi & r^2\cos^2\theta \end{bmatrix} r^2\sin\theta d\theta d\phi dr$
  $= \begin{bmatrix} \frac{4\pi}{15} & 0 & 0 \\ 0 & \frac{4\pi}{15} & 0 \\ 0 & 0 & \frac{2}{15} \end{bmatrix}$,

- $c_1(\kappa) = \frac{\Gamma(p/2)}{(2\pi)^{p/2}} M(1/2, p/2, \kappa)^{-1}$,

- $c_1(2\kappa) = \frac{\Gamma(p/2)}{(2\pi)^{p/2}} M(1/2, p/2, 2\kappa)^{-1}$.

- $\frac{\partial}{\partial\mu_1}\mathbb{E}(\boldsymbol{x}) = \int x_i(2\mu_1 {x_1}^2 + 2\mu_2 x_1 x_2 + 2\mu_3 x_1 x_3) C_w e^{\kappa(\mu^\top x)^2} dx = 0$
  $\Rightarrow 2\mu_1 \int x_i {x_1}^2 C_w e^{\kappa(\mu^\top x)^2} dx + 2\mu_2 \int x_i x_1 x_2 C_w e^{\kappa(\mu^\top x)^2} dx$
  $+ 2\mu_3 \int x_i x_1 x_3 C_w e^{\kappa(\mu^\top x)^2} dx = 0$,

- $\frac{\partial}{\partial\mu_2}\mathbb{E}(\boldsymbol{x}) = \int x_i(2\mu_2 {x_2}^2 + 2\mu_1 x_1 x_2 + 2\mu_3 x_2 x_3) C_w e^{\kappa(\mu^\top x)^2} dx = 0$
  $\Rightarrow 2\mu_2 \int x_i {x_2}^2 C_w e^{\kappa(\mu^\top x)^2} dx + 2\mu_1 \int x_i x_1 x_2 C_w e^{\kappa(\mu^\top x)^2} dx$
  $+ 2\mu_3 \int x_i x_2 x_3 C_w e^{\kappa(\mu^\top x)^2} dx = 0$,

- $\frac{\partial}{\partial\mu_3}\mathbb{E}(\boldsymbol{x}) = \int x_i(2\mu_3 {x_3}^2 + 2\mu_1 x_1 x_3 + 2\mu_2 x_2 x_3) C_w e^{\kappa(\mu^\top x)^2} dx = 0$
  $\Rightarrow 2\mu_3 \int x_i {x_3}^2 C_w e^{\kappa(\mu^\top x)^2} dx + 2\mu_1 \int x_i x_1 x_3 C_w e^{\kappa(\mu^\top x)^2} dx$
  $+ 2\mu_2 \int x_i x_2 x_3 C_w e^{\kappa(\mu^\top x)^2} dx = 0$.

To simplify the above terms we use the following notation

- $ai11 = 2\mu_1 \int x_i {x_1}^2 C_w e^{\kappa(\mu^\top x)^2} dx$,

- $ai12 = 2\mu_2 \int x_i x_1 x_2 C_w e^{\kappa(\mu^\top x)^2} dx$,

- $ai13 = 2\mu_3 \int x_i x_1 x_3 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai21 = 2\mu_1 \int x_i x_1 x_2 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai22 = 2\mu_2 \int x_i {x_2}^2 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai23 = 2\mu_3 \int x_i x_2 x_3 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai31 = 2\mu_1 \int x_i x_1 x_2 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai32 = 2\mu_2 \int x_i x_2 x_3 C_w e^{\kappa(\mu^\top x)^2} dx,$
- $ai33 = 2\mu_3 \int x_i {x_3}^2 C_w e^{\kappa(\mu^\top x)^2} dx.$

To find the third moment of the Watson distribution we must solve this system which is equivalent to solving each of these integrals in the sphere given by

$$2\begin{bmatrix} ai11 & ai12 & ai13 \\ ai21 & ai22 & ai23 \\ ai31 & ai32 & ai33 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

.

Note that $ai11 + ai22 + ai33 = 0$ and besides that $ai11 = 0$, because if $\mu = (1, 0, \cdots, 0)^\top \Rightarrow \mathbb{E}(x_i {x_1}^2) = 0$ for all $i = 1, 2, 3$. It follows, therefore, that $ai22 = -ai33$, $ai12 = \frac{-\mu_3}{\mu_2} ai13$ and $ai23 = \frac{-\left[\frac{\mu_1\mu_2^2}{\mu_3} - \mu_1\mu_3\right] ai12}{\mu_3^2 {\mu_2}^2}$. Then after some algorithms we find $ai12 = 0$, as $ai12 = ai21 = 0$. Therefore it follows that $ai23 = ai32 = 0$. Finally $ai33 = 0$ and $ai22 = 0$ as we wanted to demonstrate.

Assume that $\boldsymbol{X} \sim W(\mu, \kappa)$, $M(\mu, \kappa) := \mathbb{E}(\boldsymbol{x}\boldsymbol{x}\top)$, matrix $p \times p$. Like this $\int_{\mathbb{S}^2} \boldsymbol{x}\boldsymbol{x}\top C(\kappa) e^{\kappa(\mu^\top \boldsymbol{x})^2} d\boldsymbol{x} = M(\mu, \kappa)$. Admit simplification $(\mu_1, \mu_2, \mu_3)^\top$. So

$$\frac{\partial}{\partial \mu_i} \int_{\mathbb{S}^2} \boldsymbol{x}\boldsymbol{x}\top C(\kappa) e^{\kappa(\mu^\top \boldsymbol{x})^2} d\boldsymbol{x} = \frac{\partial}{\partial \mu_i} M(\mu, \kappa) := \boldsymbol{D}_i \tag{C.1}$$

to $i = 1, 2, 3$. As $(\boldsymbol{\mu}^\top \boldsymbol{x})^2 = \sum_{i=1}^3 \mu_i x_i + 2\sum_{1 \le i < j \le 3} \mu_i \mu_j x_i x_j$, from (14), the following system is valid

$$\begin{cases} 2\mu_1 \mathbb{E}(x_1^2 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_2 \mathbb{E}(x_1 x_2 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_3 \mathbb{E}(x_1 x_3 \boldsymbol{x}\boldsymbol{x}\top) = \boldsymbol{D1} \\ 2\mu_2 \mathbb{E}(x_2^2 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_1 \mathbb{E}(x_1 x_2 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_3 \mathbb{E}(x_2 x_3 \boldsymbol{x}\boldsymbol{x}\top) = \boldsymbol{D2} \\ 2\mu_3 \mathbb{E}(x_3^2 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_1 \mathbb{E}(x_1 x_3 \boldsymbol{x}\boldsymbol{x}\top) + 2\mu_2 \mathbb{E}(x_2 x_3 \boldsymbol{x}\boldsymbol{x}\top) = \boldsymbol{D3} \end{cases}$$

In addition to this system, the following assumption is valid: As $\sum_{i=1}^3 x_i^2 = 1$ once

$$\left\| \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right\|^2 = \sum_{i=1}^{3} {x_i}^2 = 1$$

then $M(\mu\kappa) = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) = \mathbb{E}\left[(\sum_{i=1}^{3} x_i^2)\boldsymbol{x}\boldsymbol{x}^\top\right] = \mathbb{E}(x_1^2\boldsymbol{x}\boldsymbol{x}\top) + \mathbb{E}(x_2^2\boldsymbol{x}\boldsymbol{x}\top) + \mathbb{E}(x_3^2\boldsymbol{x}\boldsymbol{x}\top)$. Defined the arrays 3×3, $\boldsymbol{Aii} = \mathbb{E}(x_i^2\boldsymbol{x}\boldsymbol{x}\top)$ and $\boldsymbol{Aij} = \mathbb{E}(x_i x_j\boldsymbol{x}\boldsymbol{x}\top)$.

Note that

$$\boldsymbol{A12} = a1\boldsymbol{A33} + \boldsymbol{b1},$$

$$\boldsymbol{A13} = a2\boldsymbol{A22} + \boldsymbol{b2},$$

$$\boldsymbol{A23} = a3\boldsymbol{A11} + \boldsymbol{b3},$$

on what $a1 = \frac{{\mu_3}^2}{\mu_1\mu_2}$, $a2 = \frac{{\mu_2}^2}{\mu_1\mu_3}$ and $a3 = \frac{{\mu_1}^2}{\mu_2\mu_3}$. We used the notation $\boldsymbol{b1} = \frac{\mu_1\boldsymbol{D1}+\mu_2\boldsymbol{D2}-\mu_3\boldsymbol{D3}-M(\mu,\kappa)}{4\mu_1\mu_2}$, $\boldsymbol{b2} = \frac{\mu_1\boldsymbol{D1}+\mu_3\boldsymbol{D3}-\mu_2\boldsymbol{D2}-M(\mu,\kappa)}{4\mu_1\mu_3}$, $\boldsymbol{b3} = \frac{\mu_2\boldsymbol{D2}+\mu_3\boldsymbol{D3}-\mu_1\boldsymbol{D1}-M(\mu,\kappa)}{4\mu_2\mu_3}$.

APPENDIX D – DERIVATION OF THE THEOREM 4.1

Let $x \sim W_p(\mu_1, \kappa_1)$ and $x \sim W_p(\mu_2, \kappa_2)$ having densities $f(z; \mu_1, \kappa_1)$ and $f(z; \mu_2, \kappa_2)$, respectively.

$$\begin{aligned}
\boldsymbol{K}_\beta(\kappa_1, \mu_1, \kappa_2, \mu_2) &\equiv \int_{\mathbb{S}^{p-1}} f^\beta(z; \mu_1, \kappa_1) f^{1-\beta}(z; \mu_2, \kappa_2) dz \\
&= \frac{(2\pi^{p/2})^{-1}\Gamma\left(\frac{p}{2}\right)}{M^\beta(\frac{1}{2}, \frac{p}{2}, \kappa_1) M^{1-\beta}(\frac{1}{2}, \frac{p}{2}, \kappa_2)} \int_{\mathbb{S}^{p-1}} \exp\{\beta\kappa_1({\mu_1}^\top z)^2 + (1-\beta)\kappa_2({\mu_2}^\top z)^2\} dz.
\end{aligned}$$

Note that the following quadratic form representation holds: For $\beta \in (0, 1)$,

$$\beta\kappa_1({\mu_1}^\top z)^2 + (1-\beta)\kappa_2({\mu_2}^\top z)^2 = \boldsymbol{z}^\top[\underbrace{\beta\kappa_1\mu_1{\mu_1}^\top + (1-\beta)\kappa_2\mu_2{\mu_2}^\top}_{A}]\boldsymbol{z},$$

which is either positive-definite or negative-definite one for $\kappa_1, \kappa_2 > 0$ (bipolar forms) or $\kappa_1, \kappa_2 < 0$ (girdle forms), respectively. According to Dryden (2005, 1655), the Watson model is a special case of the Bingham distribution pioneered by Christopher Bingham (1974) given by

$$f(z; \Sigma) = C^{-1}(\Sigma)\exp(-\boldsymbol{z}\Sigma\boldsymbol{z}) \tag{D.1}$$

for $\boldsymbol{z} \in \boldsymbol{S}^{p-1}$. The following result is valid for the normalizing constant $C(\Sigma)$ (see Kume and Wood 2005, Eq. (3)): Setting $\lambda_1, \cdots, \lambda_p$ as eigenvalues of Σ,

$$C(\Sigma) = \int_{\boldsymbol{S}^{p-1}} \exp\left(-\sum_{i=1}^{p} \lambda_i \boldsymbol{z}_i^2\right) d\boldsymbol{z}.$$

Thus, defining $\lambda_i(\kappa_1, \mu_1, \kappa_2, \mu_2, \beta)$ as ith eigenvalue of $-A$,

$$\boldsymbol{K}_\beta(\kappa_1, \mu_1, \kappa_2, \mu_2) = \frac{(2\pi^{p/2})^{-1}\Gamma\left(\frac{p}{2}\right)\int_{\boldsymbol{S}^{p-1}} \exp\left[-\sum_{i=1}^{p} \lambda_i(\kappa_1, \mu_1, \kappa_2, \mu_2, \beta)\boldsymbol{z}_i^2\right] d\boldsymbol{z}}{M^\beta(\frac{1}{2}, \frac{p}{2}, \kappa_1) M^{1-\beta}(\frac{1}{2}, \frac{p}{2}, \kappa_2)}.$$

APPENDIX E – DERIVATION OF THE COROLLARY 4.1

According to Dryden (2005, 1655), assuming $\Sigma^{-1} = \mathbb{I}_p - 2\kappa\mu\mu^\top$, the Bingham model collapses in $W_p(\mu, \kappa)$ and, therefore,

$$C(\Sigma) = \int_{\boldsymbol{S}^{p-1}} \exp\left[\kappa(\mu^\top \boldsymbol{z})^2\right] d\boldsymbol{z} = 2\pi^{p/2}\Gamma^{-1}\left(\frac{p}{2}\right) M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right).$$

Thus $K_\beta(\kappa_1, \kappa_2) \equiv K_\beta(\kappa_1, \mu, \kappa_2, \mu)$ is given by

$$\boldsymbol{K}_\beta(\kappa_1, \kappa_2) = \frac{M(\frac{1}{2}, \frac{p}{2}, \beta\kappa_1 + (1-\beta)\kappa_2)}{M^\beta(\frac{1}{2}, \frac{p}{2}, \kappa_1)M^{1-\beta}(\frac{1}{2}, \frac{p}{2}, \kappa_2)}.$$