

Bootstrap and Permutation tests in ANOVA for directional data

Adelaide Figueiredo

Faculty of Economics of University of Porto and LIAAD-INESC TEC Porto - PORTUGAL

Abstract. The problem of testing the null hypothesis of a common direction across several populations defined on the hypersphere arises frequently when we deal with directional data. We may consider the Analysis of Variance (ANOVA) for testing such hypotheses. However, for the Watson distribution, a commonly used distribution for modeling axial data, the ANOVA test is only valid for large concentrations. So we suggest to use alternative tests, such as bootstrap and permutation tests in ANOVA. Then, we investigate the performance of these tests for data from Watson populations defined on the hypersphere.

Keywords: *Hypersphere, Monte Carlo methods, simulation, Watson distribution.*

1 Introduction

The statistical analysis of directional data, represented by points on the surface of the unit sphere in \mathbb{R}^q , denoted by $S_{q-1} = \{\mathbf{x} \in \mathbb{R}^q : \mathbf{x}'\mathbf{x} = 1\}$ was widely developed by Watson (1983), Fisher *et al.* (1987), Fisher (1993), Mardia and Jupp (2000), among other authors. The applications of directional data are essentially on the circle ($q = 2$) and on the sphere ($q = 3$), but the applications on higher dimensions ($q \geq 4$) are also relevant. 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. 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.

In this paper we focus on the ANOVA test for axial data i.e. unsigned unit vectors and we consider the bootstrap version of this test and the respective permutation test. The bootstrap test consists in resampling with replacement from each sample and a permutation test consists in resampling without replacement from the whole sample. We evaluate the performance of these tests when data come from Watson populations defined on the hypersphere. We consider the Watson distribution defined on the hypersphere because it is one of the most used distributions for modeling axial data. For this distribution, the ANOVA statistic follows an F -distribution that is appropriate only for highly concentrated data (see Stephens, 1992, Gomes and Figueiredo, 1999 and Mardia and Jupp, 2000, p. 240). Thus, it seems us that the bootstrap test and permutation test based on the ANOVA statistic may perform better when data are not sufficiently highly concentrated.

The article is organized as follows. In Section 2 we refer the Watson distribution defined on the hypersphere and we present ANOVA test for this distribution. In Section 3 we propose the bootstrap approach and the permutation test to ANOVA test. In Section 4 we present numerical results about the performance of the tests in the two-sample case and in three-sample case, such as the estimation of the levels of significance of the tests and the empirical power of the tests. In Section 5 we present an application and finally, in Section 6 we conclude the paper with some remarks.

2 Analysis of Variance for axial data from Watson distributions defined on the hypersphere

In this section we refer the Watson distribution defined on the hypersphere and the ANOVA test for this distribution.

2.1 Watson distribution

The bipolar Watson distribution defined on the q -dimensional sphere, denoted by $W_q(\mathbf{u}, \kappa)$, has probability density function given by

$$f(\pm \mathbf{x}) = \left\{ {}_1F_1\left(\frac{1}{2}, \frac{q}{2}, \kappa\right) \right\}^{-1} \exp\left\{\kappa(\mathbf{u}'\mathbf{x})^2\right\}, \quad \pm \mathbf{x} \in S_{q-1}, \quad \pm \mathbf{u} \in S_{q-1}, \quad \kappa > 0, \quad (2.1)$$

where ${}_1F_1(1/2, q/2, \kappa)$ is the confluent hypergeometric function defined by

$${}_1F_1\left(\frac{1}{2}, \frac{q}{2}, \kappa\right) = \frac{\Gamma\left(\frac{q}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{q-1}{2}\right)} \int_0^1 \exp(\kappa t) t^{-0.5} (1-t)^{(q-3)/2} dt. \quad (2.2)$$

This distribution has two parameters: a directional parameter $\pm \mathbf{u}$ and a concentration parameter κ , which measures the concentration around $\pm \mathbf{u}$. It is rotationally symmetric about the principal axis $\pm \mathbf{u}$.

The Watson distribution $W_q(\mathbf{u}, \kappa)$ has the following property given in Mardia and Jupp (2000 p. 236):

For $\pm \mathbf{x} \in S_{q-1}$ from a bipolar Watson population, we have for large κ ($\kappa \rightarrow \infty$)

$$2\kappa \left\{ 1 - (\mathbf{u}'\mathbf{x})^2 \right\} \sim \chi_{q-1}^2. \quad (2.3)$$

Let $X = [\pm \mathbf{x}_1 | \pm \mathbf{x}_2 | \dots | \pm \mathbf{x}_n]$ be a random sample of size n from the bipolar Watson distribution $W_q(\mathbf{u}, \kappa)$. The maximum likelihood estimators of the parameters, given for instance, in Mardia and Jupp (2000, p. 202) and Watson (1983, p. 183-184) are defined by

- The maximum likelihood estimator $\hat{\mathbf{u}}$ of \mathbf{u} is the eigenvector of the orientation matrix XX' associated with the largest eigenvalue \hat{w} .
- The maximum likelihood estimator $\hat{\kappa}$ of κ is the solution of $Y(\hat{\kappa}) = \hat{w}/n$, where $Y(\kappa)$ is defined by

$$Y(\kappa) = \frac{d_1 F_1\left(\frac{1}{2}, \frac{q}{2}, \kappa\right)}{d\kappa}. \quad (2.4)$$

2.2 ANOVA test for Watson distribution

Let $X_i = [\pm \mathbf{x}_{i1} | \pm \mathbf{x}_{i2} | \dots | \pm \mathbf{x}_{in_i}]$, $i = 1, \dots, k$ be k independent random samples of sizes n_1, \dots, n_k from Watson distributions $W_q(\pm \mathbf{u}_i, \kappa_i)$ with polar axis $\pm \mathbf{u}_i$ and concentration parameter κ_i around $\pm \mathbf{u}_i$, $i = 1, \dots, k$ and let $n = n_1 + \dots + n_k$ be the total sample size.

Suppose that we wish to test

$$H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 = \dots = \pm \mathbf{u}_k = \pm \mathbf{u}, \quad (2.5)$$

against the alternative that at least one of the equalities is not satisfied.

Next we consider κ_i known. We note that when the concentration parameters κ_i are unknown, we may replace them by their maximum likelihood estimates. The maximum likelihood estimate $\hat{\kappa}_i$ for $i = 1, \dots, k$ is the solution of the equation $Y(\hat{\kappa}_i) = \hat{w}_i/n_i$, where $Y(\cdot)$ is defined in (2.4) and \hat{w}_i is the largest eigenvalue associated with $X_i X_i'$.

Consider the following identity

$$\sum_{i=1}^k \sum_{j=1}^{n_i} \kappa_i \left\{ 1 - (\hat{\mathbf{u}}' \mathbf{x}_{ij})^2 \right\} = \sum_{i=1}^k \sum_{j=1}^{n_i} \kappa_i \left\{ 1 - (\hat{\mathbf{u}}'_i \mathbf{x}_{ij})^2 \right\} + \sum_{i=1}^k \sum_{j=1}^{n_i} \kappa_i \left\{ (\hat{\mathbf{u}}'_i \mathbf{x}_{ij})^2 - (\hat{\mathbf{u}}' \mathbf{x}_{ij})^2 \right\}, \quad (2.6)$$

where $\hat{\mathbf{u}}_i$ is the eigenvector associated with the largest eigenvalue $\hat{\lambda}_i$ of the matrix $\kappa_i X_i X_i'$, i.e.,

$$(\kappa_i X_i X_i') \hat{\mathbf{u}}_i = \hat{\lambda}_i \hat{\mathbf{u}}_i, \quad (2.7)$$

and $\hat{\mathbf{u}}$ is the eigenvector associated with the largest eigenvalue $\hat{\lambda}$ of the matrix $\sum_{i=1}^k \kappa_i X_i X_i'$, i.e.,

$$\left(\sum_{i=1}^k \kappa_i X_i X_i' \right) \hat{\mathbf{u}} = \hat{\lambda} \hat{\mathbf{u}}. \quad (2.8)$$

The identity (2.6) is the decomposition of the total variability into the sum of the within-groups variability and the between-groups variability, and it may be written as

$$\sum_{i=1}^k \kappa_i n_i - \hat{\lambda} = \left(\sum_{i=1}^k \kappa_i n_i - \sum_{i=1}^k \hat{\lambda}_i \right) + \left(\sum_{i=1}^k \hat{\lambda}_i - \hat{\lambda} \right). \quad (2.9)$$

The test statistic is defined by

$$F = \frac{\left(\sum_{i=1}^k \hat{\lambda}_i - \hat{\lambda} \right) / (k-1)(q-1)}{\sum_{i=1}^k \left(\kappa_i n_i - \hat{\lambda}_i \right) / (n-k)(q-1)}, \quad (2.10)$$

and it may be written as:

$$F = \frac{\left(\sum_{i=1}^k \hat{\mathbf{u}}_i' (\kappa_i X_i X_i') \hat{\mathbf{u}}_i - \hat{\mathbf{u}}' \sum_{i=1}^k \kappa_i X_i X_i' \hat{\mathbf{u}} \right) / (k-1)(q-1)}{\left(\sum_{i=1}^k (\kappa_i n_i - \hat{\mathbf{u}}_i' (\kappa_i X_i X_i') \hat{\mathbf{u}}_i) \right) / (n-k)(q-1)}. \quad (2.11)$$

In the particular case of all concentration parameters equal to κ (known or unknown), the statistic given by (2.10) reduces to the following statistic

$$F = \frac{\left(\sum_{i=1}^k \hat{w}_i - \hat{w} \right) / (k-1)(q-1)}{\left(n - \sum_{i=1}^k \hat{w}_i \right) / (n-k)(q-1)}, \quad (2.12)$$

where \hat{w} is the largest eigenvalue of $\sum_{i=1}^k X_i X_i'$ and \hat{w}_i is the largest eigenvalue of $X_i X_i'$.

The test statistic F has under the null hypothesis, approximately $F_{(k-1)(q-1), (n-k)(q-1)}$ distribution, for known and large concentration parameters κ_i ($\kappa_i \rightarrow \infty, i = 1, \dots, k$).

3 Bootstrap procedure and permutation test

We consider the null hypothesis of a common polar axis, $H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 = \dots = \pm \mathbf{u}_k = \pm \mathbf{u}$ for k populations with polar axis $\pm \mathbf{u}_i$ and concentration parameter κ_i around $\pm \mathbf{u}_i$. We propose the bootstrap and permutation versions of the ANOVA statistic defined by (2.11). The algorithms for performing the bootstrap and permutation tests are based on Monte Carlo sampling in both algorithms. Amaral *et al.* (2007) refer that a key point in bootstrap hypothesis testing is make a preliminary transformation of the data before performing resampling under the null hypothesis. This is because typically the data do not satisfy the null hypothesis exactly. These authors refer a method to move $\hat{\mathbf{u}}_i$ to $\hat{\mathbf{u}}$, which will be described next. Given two unit vectors \mathbf{a} and \mathbf{b} in \mathbb{R}^q , the rotation matrix to move \mathbf{b} to \mathbf{a} along the geodesic path on the unit sphere in \mathbb{R}^q that connects \mathbf{b} to \mathbf{a} is given by

$$Q = I_p + (\sin \alpha) A + \{(\cos \alpha) - 1\} (\mathbf{a}\mathbf{a}' + \mathbf{c}\mathbf{c}'), \quad (3.1)$$

where $\alpha = \cos^{-1}(\mathbf{a}'\mathbf{b}) \in (0, \pi)$ and $A = \mathbf{a}\mathbf{c}' + \mathbf{c}\mathbf{a}'$, with $\mathbf{c} = \frac{\mathbf{b} - \mathbf{a}(\mathbf{a}'\mathbf{b})}{\|\mathbf{b} - \mathbf{a}(\mathbf{a}'\mathbf{b})\|}$, where $\|\cdot\|$ denotes the Euclidean norm on \mathbb{R}^q . Then, $Q\mathbf{b} = \mathbf{a}$, in our case $\mathbf{b} = \hat{\mathbf{u}}_i$ and $\mathbf{a} = \hat{\mathbf{u}}$. The theoretical accuracy of the bootstrap procedure was analyzed in Amaral *et al.* (2007).

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

1. For each sample of size n_i , calculate the estimate of \mathbf{u}_i defined by (2.7), $\hat{\mathbf{u}}_i$ and the corresponding eigenvalue $\hat{\lambda}_i$, $i = 1, \dots, k$.
2. Determine the estimate of the common polar axis $\hat{\mathbf{u}}$, defined by (2.8), and the corresponding eigenvalue $\hat{\lambda}$. Then, calculate the statistic value F_{obs} defined in (2.10).
3. Transform each sample i using the rotation matrix (3.1) to move $\hat{\mathbf{u}}_i$ to $\hat{\mathbf{u}}$ ($i = 1, \dots, k$).
4. For each bootstrap cycle b , $b = 1, \dots, B$ do as follows. For $i = 1, \dots, k$ draw a re-sample of size n_i randomly with replacement, from the sample i , and calculate the eigenvalue $\hat{\lambda}_i^{(b)}$ using (2.7) for known concentration parameters κ_i and calculate the eigenvalue $\hat{\lambda}_i^{(b)}$ and $\hat{\kappa}_i^{(b)}$ for unknown concentration parameters. Then, determine the bootstrap statistic $F^{(b)}$ defined by

$$F^{(b)} = \frac{\left(\sum_{i=1}^k \hat{\lambda}_i^{(b)} - \hat{\lambda} \right) / (k-1)(q-1)}{\sum_{i=1}^k \left(\kappa_i n_i - \hat{\lambda}_i^{(b)} \right) / (n-k)(q-1)}. \quad (3.2)$$

and

$$F^{(b)} = \frac{\left(\sum_{i=1}^k \hat{\lambda}_i^{(b)} - \hat{\lambda} \right) / (k-1)(q-1)}{\sum_{i=1}^k \left(\hat{\kappa}_i^{(b)} n_i - \hat{\lambda}_i^{(b)} \right) / (n-k)(q-1)}, \quad (3.3)$$

for known and unknown κ_i , respectively.

5. Determine the bootstrap p-value by

$$\hat{p} = \frac{1 + \sum_{b=1}^B I_{\{F^{(b)} \geq F_{obs}\}}}{B + 1}, \quad (3.4)$$

where the indicator function is defined by $I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$

The algorithm for implementing the permutation test can be described in the following four steps. Let $[\pm \mathbf{x}_{i1} | \pm \mathbf{x}_{i2} | \dots | \pm \mathbf{x}_{in_i}]$ be the i -sample of unit vectors.

1. For each sample $i = 1, \dots, k$, calculate the eigenvalue $\hat{\lambda}_i$ defined by (2.7), and then the eigenvalue $\hat{\lambda}$ defined by (2.8).
2. Determine the statistic value F_{obs} given in (2.10).
3. At each permutation cycle c , $c = 1, \dots, C$ do as follows. Sample randomly, without replacement, from the pooled set of observations $[\pm \mathbf{x}_{i1} | \pm \mathbf{x}_{i2} | \dots | \pm \mathbf{x}_{in_i}]$, $i = 1, \dots, k$, $j = 1, \dots, n_i$ to form k subsamples of sizes n_1, \dots, n_k and for each, calculate the eigenvalue $\hat{\lambda}_i^{(c)}$ using (2.7) for known concentration parameters κ_i and for unknown concentration parameters, calculate $\hat{\lambda}_i^{(c)}$ and $\hat{\kappa}_i^{(c)}$. Next, determine the permutation version of the statistic $F^{(c)}$ defined by

$$F^{(c)} = \frac{\left(\sum_{i=1}^k \hat{\lambda}_i^{(c)} - \hat{\lambda} \right) / (k-1)(q-1)}{\sum_{i=1}^k \left(\kappa_i n_i - \hat{\lambda}_i^{(c)} \right) / (n-k)(q-1)}, \quad (3.5)$$

and

$$F^{(c)} = \frac{\left(\sum_{i=1}^k \hat{\lambda}_i^{(c)} - \hat{\lambda} \right) / (k-1)(q-1)}{\sum_{i=1}^k \left(\hat{\kappa}_i^{(c)} n_i - \hat{\lambda}_i^{(c)} \right) / (n-k)(q-1)}, \quad (3.6)$$

for known and unknown κ_i , respectively.

4. Determine the permutation p -value by

$$\hat{p} = \frac{1 + \sum_{c=1}^C I_{\{F^{(c)} \geq F_{obs}\}}}{C + 1}, \quad (3.7)$$

where $I(\cdot)$ is the indicator function previously defined.

The permutation tests in k -samples problems are in general valid if under the null hypothesis the k sets of observations are exchangeable, i.e., the k populations are identical with the same parameters (see Wellner, 1979, Romano, 1990, Good, 2004 and Amaral *et al.*,

2007). In our experimental analysis in next section, we will consider not only the case of identical populations with the same parameters under the null hypothesis, but also the case of populations with the same polar axis and different concentration parameters under the null hypothesis.

4 Performance of the tests for data from Watson populations

In this section, we present the results of the performance of the tests obtained with a simulation study. We note that in this study as we can not present all the cases studied, we selected only some cases that seemed relevant to us. For the simulation of the Watson distribution defined on the hypersphere, we used the acceptance-rejection method given in Li and Wong (1993).

First, we suppose two Watson populations with known concentrations in subsection 4.1. Second, we consider two Watson populations with estimated concentrations in subsection 4.2. Third, we suppose three Watson populations with a common and known concentration parameter in subsection 4.3.

4.1 Two Watson populations with known concentrations (equal or different)

We considered two Watson populations $W_q(\mathbf{u}_1, \kappa_1)$ and $W_q(\mathbf{u}_2, \kappa_2)$, where the concentration parameters κ_1 and κ_2 are known.

An extensive simulation study was undertaken and we present the results for the dimensions of the sphere $q = 2, 3, 4, 5$ to test $H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 = \pm \mathbf{u}$. This study was carried out for investigating the performance of the three tests for the ANOVA statistic given by (2.10), the tabular test, based directly on the null asymptotic distribution of the statistic, the bootstrap test and the permutation test. We note that for equal concentration parameters, the ANOVA statistic reduces to the statistic (2.12), which does not depend on the common concentration parameter.

First we estimated the significance level of the three tests and second, we determined the empirical power of the tests.

4.1.1 Estimated significance levels

We considered without loss of generality, that under $H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 = \pm \mathbf{e}_q$, where $\mathbf{e}_q = (0, \dots, 0, 1)'$. We generated two samples of sizes n_1 and n_2 of the populations $W_q(\mathbf{e}_q, \kappa_1)$ and $W_q(\mathbf{e}_q, \kappa_2)$, supposing samples of equal size and also samples of different sizes. The estimated levels of significance obtained for a nominal significance level of 5% are indicated

in Tables 1-2 for known and equal concentration parameters ($\kappa_1 = \kappa_2 = \kappa$) and known and different concentration parameters ($\kappa_1 \neq \kappa_2$), respectively. In these tables we highlighted in bold 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 10000 Monte Carlo simulations in the tabular test and 5000 Monte Carlo simulations in the bootstrap and permutation tests. The number of bootstrap re-samples, B , in each Monte Carlo simulation was $B = 200$ and the number of permutation samples was $C = 200$. For obtaining the significance levels, we used the 0.95-percentile of an F -distribution in the tabular test and the 0.95-percentile of the distribution of values of the bootstrap and permutation statistics in the bootstrap and permutation tests, respectively.

The estimated significance levels obtained for equal and different concentration parameters enable us to draw similar conclusions. First, we note that in the tabular test, although we used the critical point of an F -distribution for a significance level of 5%, the estimated significance levels obtained are not exactly equal to the nominal significance level of 5%. The estimated significance levels in the tabular test are in general more distant from the nominal significance level for small concentration parameters. We note that in these tables we did not consider very large values of the concentration parameters, for which the significance levels obtained in the tabular test are the closest to the nominal level of significance. This would be expected since the F -distribution of the test statistic is only an asymptotic distribution, valid for large concentrations. Thus, it is necessary to consider other versions of the ANOVA statistic such as the bootstrap and permutation versions.

Second, the significance levels obtained in the permutation test in the case of equal or different 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.

Third, 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 and has generally similar accuracy to the tabular test in the case of large concentration parameters.

Finally, we note that the estimated levels of significance of the tests for a nominal level of significance of 1% led us to similar conclusions.

4.1.2 Empirical power of the tests

Second, 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 $\pm \mathbf{u}_1$ and $\pm \mathbf{u}_2$ which form an angle θ between them, with $\theta = 18^\circ, 36^\circ, 54^\circ, 72^\circ, 90^\circ$. Thus, under this alternative

Table 1: Estimated significance levels (in %) of the tabular, bootstrap and permutation tests, for a common and known concentration parameter κ and several sample sizes n_1, n_2 .

n_1, n_2	κ	$q = 2$			$q = 3$			$q = 4$			$q = 5$		
		Tab.	Boot.	Perm.	Tab.	Boot.	Perm.	Tab.	Boot.	Perm.	Tab.	Boot.	Perm.
3, 5	1	5.2	4.6	5.7	3.1	5.3	5.3	3.7	5.8	5.6	4.3	6.2	6.2
	2	4.4	4.4	5.6	3.1	5.3	5.3	3.7	5.8	5.6	4.3	6.2	6.2
	5	3.8	3.7	4.9	3.1	5.3	5.3	3.7	5.2	5.6	4.3	6.2	6.2
	10	4.1	3.8	4.7	3.8	5.1	4.8	4.0	5.2	4.9	3.8	5.3	5.1
	20	4.6	4.0	4.7	4.7	5.1	4.8	4.3	5.3	4.8	4.5	5.3	4.9
5, 10	1	8.6	5.5	6.1	7.7	7.0	6.1	12.1	7.0	6.1	5.1	7.1	5.7
	2	6.2	5.4	6.0	6.7	6.5	6.1	9.7	6.6	6.2	5.6	6.8	5.7
	5	3.4	5.5	5.3	3.2	6.0	5.4	3.9	5.8	5.8	5.8	5.8	5.6
	10	3.7	5.5	5.1	3.7	6.4	5.2	5.8	6.3	5.2	3.7	6.3	5.0
	20	4.6	4.0	4.7	4.7	5.1	4.8	4.3	5.3	4.8	4.6	6.5	5.2
5, 5	1	6.7	5.2	5.7	3.2	5.5	5.3	3.6	6.1	5.7	4.6	6.4	6.1
	2	5.0	5.1	5.6	3.2	5.5	5.3	3.6	6.1	5.7	4.6	6.4	6.1
	5	2.8	4.9	5.9	3.2	5.5	5.3	3.6	6.1	5.7	4.6	6.4	6.1
	10	3.9	4.9	5.0	3.8	5.2	5.0	3.7	5.3	5.0	3.1	5.4	5.1
	20	4.3	4.9	5.0	4.5	5.2	4.9	3.9	5.4	4.9	4.1	5.3	5.0
10, 10	1	11.0	5.6	6.3	2.8	5.5	5.3	3.1	5.7	5.6	4.2	5.8	5.9
	2	6.5	5.7	6.1	2.8	5.5	5.3	3.1	5.7	5.6	4.2	5.8	5.9
	5	2.9	5.6	5.2	2.8	5.5	5.3	3.1	5.7	5.6	4.2	5.8	5.9
	10	4.3	5.6	5.1	4.0	5.3	5.2	3.2	5.3	5.3	2.9	5.3	5.3
	20	4.8	5.4	5.2	5.3	5.3	5.2	4.4	5.3	5.2	3.9	5.4	5.2
20, 20	1	13.5	5.8	6.5	22.3	7.4	7.0	30.6	8.0	7.0	38.7	8.0	2.6
	2	5.9	5.9	5.8	9.1	6.8	6.0	13.5	7.2	6.6	28.9	7.4	7.0
	5	2.9	5.6	5.2	2.1	6.7	5.3	2.5	6.9	5.4	2.8	6.7	5.5
	10	3.7	5.5	5.2	3.5	7.1	5.2	3.6	7.3	5.3	2.9	7.2	5.3

hypothesis and without loss of generality, we generated one sample from $W_q(\mathbf{e}_q, \kappa_1)$ and the other sample from $W_q(\mathbf{u}, \kappa_2)$, where \mathbf{u} is defined by $\mathbf{u} = (0, \dots, 0, (1 - \cos^2 \theta)^{1/2}, \cos \theta)$.

We note that if the angle θ is equal to 0° , we obtain the significance level of the tests. As in the estimation of the significance levels, to determine the empirical power of the tests, we used the 0.95-percentile of an F -distribution in the tabular test and the 0.95-percentile

Table 2: Estimated significance levels (in %) of the tabular, bootstrap and permutation tests for different and known concentration parameters κ_1 and κ_2 and several sample sizes n_1, n_2 .

n_1, n_2	κ_1, κ_2	$q = 3$			$q = 4$			$q = 5$		
		Tab.	Boot.	Perm.	Tab.	Boot.	Perm.	Tab.	Boot.	Perm.
3, 5	1,2	1.5	6.9	5.8	1.0	7.1	5.8	0.6	7.0	5.7
	3,5	3.4	5.9	5.5	2.7	6.2	5.8	2.1	6.4	5.9
	5,10	5.0	5.1	5.0	5.6	5.5	5.2	5.0	5.7	5.4
	10,20	4.5	5.1	4.8	4.8	5.2	4.9	4.6	5.4	4.9
5, 10	1, 2	5.6	5.8	6.2	5.1	5.7	6.3	4.4	5.6	6.2
	3, 5	5.6	5.8	5.6	5.1	5.7	5.9	4.4	5.6	6.2
	5, 10	4.5	5.5	5.3	5.2	5.7	5.6	5.9	5.7	5.4
	10, 20	4.6	5.4	5.2	4.9	5.3	5.2	4.4	5.4	5.3
5, 5	1,2	5.0	6.6	5.8	4.6	6.7	5.7	4.1	6.8	5.7
	3, 5	4.1	5.7	5.6	4.2	5.9	5.9	4.3	6.0	6.0
	5,10	3.7	5.6	5.1	4.4	5.6	5.3	4.6	5.9	5.4
	10,20	4.3	5.2	5.0	3.7	5.4	5.1	3.6	5.3	5.2
10, 10	1,2	9.5	6.6	6.5	11.3	6.9	6.4	12.8	6.9	6.2
	3,5	4.1	6.3	5.6	4.6	6.2	5.9	5.7	6.2	6.2
	5,10	3.2	6.5	5.3	3.5	6.4	5.4	3.8	6.4	5.4
	10,20	4.2	5.3	5.2	4.1	5.4	5.2	3.9	5.3	5.3
20, 20	1,2	12.3	14.3	6.6	17.3	8.8	6.9	24.1	13.2	6.8
	3,5	3.4	7.0	5.3	3.6	7.2	5.5	4.4	7.3	5.7
	5,10	3.3	7.0	5.3	3.3	7.0	5.3	3.1	6.9	5.4

of the bootstrap or permutation distribution in the bootstrap or permutation tests.

In the tabular test, the empirical power was obtained from 10000 replicates of the test statistic under the alternative hypothesis. In the bootstrap or permutation test, the empirical power was obtained from 5000 Monte Carlo simulations, where in each simulation, two samples were generated under H_1 and 200 bootstrap or permutation re-samples were considered.

We indicate the empirical power of the tests for equal and known concentration parameters in Table 3 for $q = 2$, in Table 4 for $q = 3$ and in Table 5 for $q = 4, 5$. In these tables we highlighted the values of the power, in which the bootstrap test is more powerful than the tabular and permutation tests. Additionally, we show in Figure 1 the empirical power of the three tests for equal and known concentration parameters when $q = 4, 5$ and $n_1 = 5, n_2 = 10$.

Table 3: Empirical power (in %) of tabular, bootstrap and permutation tests, for $q = 2$, common and known concentration κ , angle $\theta(^{\circ})$ and sample sizes n_1, n_2 .

n_1, n_2	$\theta \backslash \kappa$	Tabular				Bootstrap				Permutation			
		1	2	5	10	1	2	5	10	1	2	5	10
3, 5	18	5.8	6.1	9.5	25.4	19.0	17.9	16.7	26.0	5.7	5.7	5.3	4.9
	36	8.4	11.8	30.5	72.9	22.2	24.2	35.2	62.0	5.7	5.7	5.3	4.9
	54	15.5	24.6	61.9	97.5	30.0	36.9	59.5	91.6	5.5	5.8	5.7	3.8
	72	31.6	44.8	83.9	99.7	44.7	54.6	79.5	98.3	4.5	4.9	4.2	1.9
	90	57.7	66.9	90.2	99.8	68.2	72.6	84.4	94.2	1.8	2.2	2.2	1.8
5, 10	18	10.0	8.7	15.3	44.1	21.2	20.8	27.4	50.1	6.0	6.9	5.4	5.2
	36	16.9	22.6	55.7	96.3	26.9	32.6	59.7	91.6	5.8	6.1	5.9	5.1
	54	31.9	47.1	88.9	100.0	39.0	50.8	83.9	99.7	5.1	5.7	6.1	4.5
	72	54.2	73.1	97.3	100.0	58.1	70.3	94.4	100.0	3.3	4.2	4.8	2.3
	90	76.0	85.4	98.6	100.0	76.8	84.1	96.9	100.0	8.2	1.3	2.7	1.8
5, 5	18	7.4	7.2	10.9	32.4	20.4	19.9	23.2	38.1	5.7	5.8	5.2	5.1
	36	11.5	15.7	39.5	86.4	24.4	27.2	46.8	79.8	5.8	6.1	6.2	5.8
	54	20.1	32.8	76.1	99.8	32.2	40.5	73.4	97.8	5.9	7.1	8.3	9.1
	72	39.9	56.4	92.9	100.0	47.5	58.1	88.2	99.8	5.6	7.9	12.6	17.2
	90	59.3	73.2	96.3	100.0	64.4	73.3	92.8	99.8	3.6	6.1	14.0	19.8
10, 10	18	13.8	11.9	22.2	62.7	20.8	20.8	33.6	61.6	6.4	6.3	5.5	5.2
	36	23.3	31.2	73.1	99.7	27.4	34.9	71.5	97.6	6.6	7.2	6.7	5.9
	54	43.6	63.8	97.8	100.0	40.8	56.1	93.3	100.0	6.6	8.6	9.8	8.5
	72	68.0	85.8	99.9	100.0	60.5	76.8	98.3	100.0	5.5	9.4	15.2	15.9
	90	80.5	92.1	99.9	100.0	75.4	86.2	99.0	100.0	1.6	4.5	17.2	22.8

We indicate the empirical power of the tests for different and known concentration parameters in Table 6 for $q = 3$ and in Table 7 for $q = 4$. As before, in these tables we highlighted the values of the power, in which the bootstrap test is more powerful than the tabular and permutation tests. Figure 2 shows the empirical power of the tests for different and known concentration parameters when $q = 5$.

The results obtained with a common concentration parameter and with different concentration parameters are similar. For both cases of equal and different concentration parameters and for each dimension of the sphere q , we conclude from the three tests, that the permutation test is the one that is least powerful. For each q , despite the significance level of the permutation test be very close to the nominal level of significance in many cases,

Table 4: Empirical power (in %) of tabular, bootstrap and permutation tests, for $q = 3$, common and known concentration κ , angle $\theta(^{\circ})$ and sample sizes n_1, n_2 .

n_1, n_2	$\theta \backslash \kappa$	Tabular				Bootstrap				Permutation			
		1	2	5	10	1	2	5	10	1	2	5	10
3, 5	18	6.1	6.1	7.2	20.4	37.8	31.7	21.9	27.0	5.6	5.6	5.3	4.9
	36	7.7	9.7	20.6	68.2	39.9	36.2	34.6	61.2	5.7	5.8	5.9	5.4
	54	11.1	17.6	45.7	96.2	44.2	46.0	54.6	90.3	5.6	5.9	6.6	5.1
	72	18.3	30.1	69.5	99.7	53.5	57.8	72.2	98.1	5.8	6.0	6.8	3.5
	90	25.3	38.3	77.4	99.8	60.0	65.2	79.3	99.3	6.1	5.9	6.8	3.0
5, 10	18	8.9	8.8	10.4	37.9	29.9	24.5	18.8	35.0	6.2	6.2	5.7	5.3
	36	11.3	12.6	40.2	94.4	32.8	31.7	40.9	83.5	6.0	6.3	6.3	5.8
	54	15.5	21.8	77.3	100.0	40.2	45.4	68.2	98.6	6.1	6.6	7.3	6.2
	72	23.9	37.3	92.0	100.0	51.2	60.3	85.8	99.9	6.4	6.6	7.0	5.3
	90	31.3	45.0	95.3	100.0	58.4	67.9	89.8	100.0	7.2	7.1	6.5	4.9
5, 5	18	8.7	7.5	8.2	27.2	32.8	27.1	19.3	30.0	5.8	5.9	5.6	5.2
	36	11.4	13.1	28.7	82.7	35.1	32.3	35.5	71.9	5.8	6.2	6.9	6.6
	54	16.9	25.2	61.9	99.4	40.5	42.3	58.3	95.3	5.9	6.7	9.4	11.7
	72	26.3	43.0	85.3	100.0	49.7	56.2	76.8	99.5	5.9	7.2	13.7	18.7
	90	33.8	53.6	90.2	100.0	56.5	64.4	83.9	99.7	5.9	7.5	15.7	20.5
10, 10	18	17.9	13.7	15.2	55.2	24.2	19.3	19.7	45.2	6.5	6.7	5.8	5.5
	36	23.6	27.4	60.0	99.5	28.3	28.6	49.1	93.8	6.6	7.2	7.4	6.9
	54	36.7	54.8	93.9	100.0	36.7	45.3	80.8	99.9	6.8	8.8	10.9	10.8
	72	53.3	77.1	99.2	100.0	50.3	64.7	94.5	100.0	6.8	10.1	17.7	18.3
	90	61.4	83.3	99.5	100.0	58.1	72.1	96.5	100.0	6.7	10.1	23.2	35.0
20, 20	18	24.9	18.4	31.1	88.0	17.3	15.4	29.1	70.4	7.1	6.4	5.7	5.4
	36	37.8	49.2	92.5	100.0	23.0	30.8	77.2	99.8	7.3	7.6	7.3	6.8
	54	62.8	85.1	100.0	100.0	37.1	59.3	97.6	100.0	7.8	10.1	11.1	10.7
	72	79.3	95.6	100.0	100.0	55.2	80.6	99.7	100.0	8.1	12.8	18.4	20.2
	90	83.7	97.3	100.0	100.0	62.7	86.4	99.8	100.0	8.0	12.6	23.6	40.4

the empirical power of this test remains close to the significance level for low values of the concentration parameters (equal or not) or when the sample sizes are different. For large values of the concentration parameters (equal or not) and for each q , the permutation test has better performance for equal-sized samples than for samples of different sizes, since the empirical power increases as the angle increases for equal-sized samples, while it remains

Table 5: Empirical power (in %) of tabular, bootstrap and permutation tests, for $q = 4, 5$, common and known concentration κ , angle $\theta(^{\circ})$ and sample sizes n_1, n_2 .

n_1, n_2	q	$\theta \backslash \kappa$	Tabular				Bootstrap				Permutation			
			1	2	5	10	1	2	5	10	1	2	5	10
3, 5	4	18	7.6	7.2	7.0	16.6	46.0	40.5	26.0	23.1	5.5	5.6	5.8	5.1
		36	12.7	14.5	20.3	50.6	51.7	48.8	39.4	48.0	5.7	5.7	6.3	5.9
		54	18.9	23.4	41.0	83.4	58.1	58.1	55.5	73.1	5.7	5.5	7.0	7.1
		72	22.6	32.8	59.2	97.5	61.8	64.0	68.3	89.6	5.6	5.6	7.3	8.5
		90	25.2	35.7	65.1	99.1	63.9	66.5	72.6	95.1	5.6	5.5	7.7	9.4
	5	18	5.8	5.6	4.3	3.8	35.0	30.6	18.1	25.9	5.5	5.6	5.8	5.1
		36	5.7	6.5	5.6	14.0	36.6	34.8	29.4	69.8	5.7	5.7	6.3	5.9
		54	6.3	7.5	10.9	25.6	40.0	41.8	48.9	95.6	5.7	5.5	7.0	7.1
		72	10.5	16.6	43.9	98.5	45.5	50.6	67.1	99.4	5.6	5.6	7.3	8.5
		90	11.5	19.3	50.9	99.1	47.9	53.9	73.0	99.7	5.6	5.5	7.7	9.4
5, 5	4	18	11.4	11.0	8.8	20.3	39.3	33.8	22.1	25.2	5.8	5.9	6.1	5.4
		36	17.9	21.4	28.7	66.4	44.9	42.8	36.2	57.7	5.7	6.1	7.5	7.0
		54	25.6	34.9	57.7	94.3	51.4	53.2	55.8	84.1	5.6	6.1	10.1	11.2
		72	31.6	45.9	77.7	99.8	56.5	59.9	70.1	96.4	5.4	6.4	13.4	17.9
		90	34.1	49.6	83.0	100.0	57.7	62.5	75.8	99.1	5.6	6.5	14.9	21.8
	5	18	9.0	8.8	7.0	17.3	39.9	34.7	21.4	23.2	5.9	6.1	6.4	5.3
		36	10.5	11.7	17.0	70.3	40.9	36.9	29.5	57.8	5.8	6.3	7.3	7.1
		54	13.4	17.5	39.0	97.5	43.1	42.9	43.4	87.8	5.9	6.4	9.2	12.4
		72	15.9	26.4	64.2	99.9	47.0	49.7	59.8	97.8	5.9	6.5	12.4	19.0
		90	18.9	31.5	72.7	100.0	49.4	53.6	66.3	99.1	6.1	6.7	14.3	23.1
10, 10	4	18	27.1	22.3	15.0	41.6	30.8	25.1	19.2	36.6	6.0	6.4	6.2	5.6
		36	41.5	46.9	61.3	95.6	40.3	39.0	48.0	82.8	5.5	6.5	8.2	7.3
		54	54.0	68.3	92.1	100.0	49.3	55.2	76.7	98.4	4.9	6.6	12.4	11.3
		72	62.4	79.5	98.0	100.0	56.4	65.3	89.6	100.0	4.5	6.8	18.6	20.4
		90	64.5	82.3	98.7	100.0	57.3	68.7	92.3	100.0	4.4	6.8	22.2	39.2
	5	18	22.3	18.5	9.7	39.2	28.4	23.4	15.3	34.0	6.4	6.9	6.4	5.6
		36	25.1	26.7	38.4	97.7	30.4	27.9	30.5	86.5	6.4	7.2	8.0	7.4
		54	31.6	43.5	78.6	100.0	34.1	37.0	57.3	99.5	6.6	8.0	11.8	12.2
		72	41.6	60.6	95.1	100.0	40.3	49.2	79.4	100.0	6.7	8.6	18.9	21.6
		90	44.8	67.1	97.0	100.0	43.6	54.0	84.8	100.0	6.7	8.8	23.8	40.5

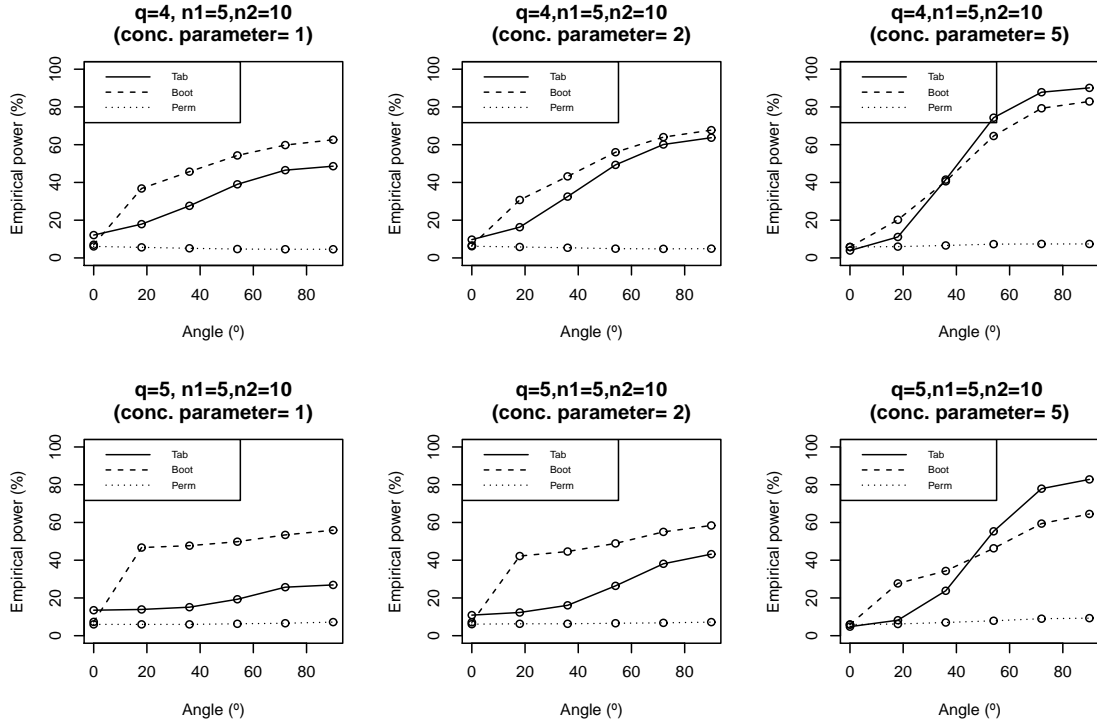


Figure 1: Empirical power of the tests for a common and known concentration

approximately equal to the nominal level of significance for different-sized samples. For each q , we also observed that for samples of equal size, the empirical power of the permutation test increases as the concentration parameters (equal or not) increase. For each q and large values of the concentration parameters (equal or not), the empirical power of the permutation test increases in general, when the common sample sizes increases.

For each dimension of the sphere q , the empirical power of the tabular and bootstrap tests increases rapidly and in some cases tends quickly to 1 when the angle between directional parameters θ increases for equal or different concentration parameters and samples of equal or different sizes. For each q , the empirical power of the tabular and bootstrap tests increases in general when the concentration parameters (equal or not) increase. For equal or different concentration parameters and for each q , the empirical power of the tabular and bootstrap tests increases in general as the sample sizes (equal or not) increase.

In both cases of equal and different concentration parameters, for each dimension of the sphere q and samples of sizes equal or not, the bootstrap test is generally more powerful than the tabular test for small concentration parameters or small angle θ between the directional parameters. The superiority of the bootstrap test compared to the tabular test is more pronounced for small samples than for large samples. For instance, is greater for $n = 5$ than for $n = 10$, and for $n_1 = 3, n_2 = 5$ than for $n_1 = 5$ and $n_2 = 10$. In fact, for

Table 6: Empirical power (in %) of tabular, bootstrap and permutation tests, for $q = 3$, different and known concentrations κ_1, κ_2 , angle $\theta(^{\circ})$ and sample sizes n_1, n_2 .

n_1, n_2	$\theta \backslash \kappa_1, \kappa_2$	Tabular			Bootstrap			Permutation		
		1,2	3,5	5,10	1,2	3,5	5,10	1,2	3,5	5,10
3, 5	18	1.9	4.9	11.1	36.6	27.7	24.9	5.8	5.5	5.1
	36	2.9	11.6	33.4	40.2	36.3	44.6	5.8	5.7	5.1
	54	5.4	24.1	63.0	46.7	50.9	66.9	6.0	6.0	4.9
	72	9.6	40.9	79.3	55.0	65.3	80.9	6.0	6.0	4.7
	90	13.5	47.8	84.0	61.4	70.5	85.0	6.3	5.6	4.8
5, 10	18	3.9	6.9	16.8	30.8	21.9	22.8	6.3	5.6	5.3
	36	6.3	21.3	58.2	35.2	36.8	52.3	6.2	6.0	5.3
	54	11.4	48.4	87.7	44.1	56.8	79.1	6.1	6.4	5.1
	72	21.4	69.7	95.4	55.1	72.9	90.0	6.4	6.0	4.3
	90	26.9	76.0	97.1	61.5	78.6	92.5	6.9	5.5	3.8
5, 5	18	5.4	7.0	13.9	31.8	22.7	23.0	5.9	5.7	5.4
	36	8.4	18.6	45.2	35.5	33.6	47.3	5.9	6.7	6.5
	54	13.0	38.8	78.6	42.1	50.3	72.7	5.9	8.0	9.3
	72	21.4	61.1	91.3	50.5	65.4	85.1	6.0	9.9	13.0
	90	26.7	68.1	93.5	56.6	71.8	88.4	6.1	10.6	15.4
10, 10	18	12.0	11.2	26.5	23.2	18.3	27.5	6.5	6.0	5.6
	36	18.7	39.5	81.0	28.9	37.4	66.4	6.7	7.0	6.8
	54	31.7	75.4	97.8	39.5	63.5	90.5	7.0	9.5	9.7
	72	47.2	90.4	99.5	51.2	80.2	96.6	7.2	12.5	14.0
	90	54.3	92.8	99.8	58.4	84.9	97.6	7.0	13.2	17.6
20, 20	18	17.0	20.3	51.1	16.6	21.4	41.6	6.7	5.7	5.7
	36	32.8	72.2	98.4	25.1	57.1	89.6	7.1	7.2	7.0
	54	58.2	96.0	100.0	40.3	85.9	99.0	7.8	9.7	9.7
	72	75.1	99.1	100.0	56.9	94.8	99.8	8.0	11.9	12.9
	90	79.8	99.5	100.0	63.7	96.4	99.9	7.9	11.1	13.4

small samples (of equal size or not), the bootstrap test is better than the tabular test for small values of the concentration parameters (equal or not) and also for large concentration parameters and small angle between the directional parameters. Thus, the results of the power indicate that the bootstrap test may be a good alternative to the tabular test for small concentration parameters (equal or not) and small samples (of equal size or not) or

Table 7: Empirical power (in %) of tabular, bootstrap and permutation tests, for $q = 4$, different and known concentrations κ_1, κ_2 , angle $\theta(^{\circ})$ and sample sizes n_1, n_2 .

n_1, n_2	$\theta \backslash \kappa_1, \kappa_2$	Tabular			Bootstrap			Permutation		
		1,2	3,5	5,10	1,2	3,5	5,10	1,2	3,5	5,10
3, 5	18	1.7	3.9	9.4	45.8	31.9	26.5	5.8	5.9	5.3
	36	3.7	10.1	21.3	51.5	41.8	37.2	5.7	6.0	5.6
	54	6.4	19.2	39.0	58.3	54.0	50.1	6.0	6.4	6.5
	72	8.7	31.6	56.5	62.7	63.7	61.7	6.1	6.6	7.0
	90	10.4	34.6	60.8	65.4	67.0	66.8	6.1	6.5	7.4
5, 10	18	4.0	6.5	11.8	36.6	24.1	20.3	5.9	6.1	5.4
	36	8.0	19.8	37.2	45.7	38.4	36.2	5.4	6.3	5.7
	54	15.4	42.8	68.3	54.8	57.1	56.7	4.9	6.3	6.0
	72	21.8	60.5	84.2	60.2	68.7	72.9	4.8	6.2	5.8
	90	23.5	65.4	88.0	63.1	72.9	78.3	4.9	6.1	5.1
5, 5	18	6.2	6.9	9.6	37.7	25.3	21.1	5.6	6.1	5.6
	36	10.3	18.1	30.0	44.2	36.2	35.8	5.7	6.9	7.2
	54	15.8	36.8	58.2	50.4	50.5	55.4	5.7	8.1	10.2
	72	21.1	53.3	77.2	55.4	61.1	70.0	5.7	9.3	14.0
	90	23.5	58.3	82.7	57.2	66.1	75.4	5.6	9.9	16.2
10, 10	18	16.3	12.0	16.9	27.9	19.7	26.0	6.0	6.3	5.8
	36	26.8	41.7	61.3	37.0	38.8	48.1	5.6	7.7	7.5
	54	39.8	73.6	91.1	46.6	61.1	75.7	5.3	9.9	10.7
	72	49.8	87.7	97.5	53.4	75.5	89.2	5.1	12.0	15.5
	90	52.6	90.2	98.5	56.6	79.2	92.7	4.9	12.8	18.5

when the alternative hypothesis is not far away from the null hypothesis.

4.2 Two Watson populations with estimated concentrations

Next, we study the effect in the performance of the tests of the estimation of the concentration parameters based on ANOVA statistic defined by (2.10). First we determined the estimated levels of significance of the tests and second, the empirical power of the tests. We considered the same number of Monte Carlo simulations in the tests as before. The number of bootstrap or permutation samples was also the same as before. The estimated levels of significance are indicated in Table 8 for $q = 3, 4$ in the case of equal or different concentration parameters. The empirical power of the tests obtained with concentration

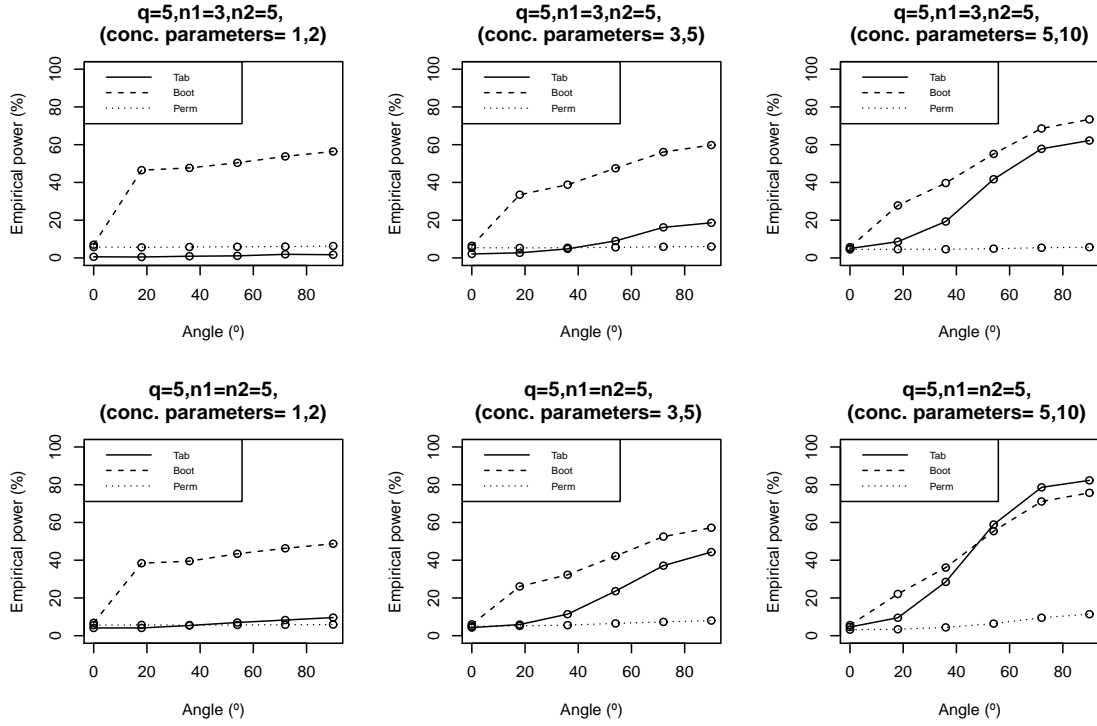


Figure 2: Empirical power of the tests for different and known concentrations

estimates for $q = 3, 4$, with a nominal significance level of 5% is indicated in Table 9 for equal concentration parameters and in Table 10 for different concentration parameters. In these tables we highlight the values of the power, in which the bootstrap test is the most powerful. The empirical power obtained for $q = 3, 4$ and $n_1 = 5, n_2 = 10$ when we estimate the concentration parameters can be seen in Figures 3 and 4 for equal and different concentration parameters, respectively.

As we may observe from these tables and figures, the results obtained for the estimated level of significance and the empirical power are not substantially affected by the estimation of the concentration parameters through the maximum likelihood method. When the concentrations are estimated, the behavior of the tests is similar to the case when the concentrations are known. Consequently, we have the same conclusions for both cases of known and estimated concentration parameters.

4.3 Three Watson populations with a common and known concentration parameter

We considered three Watson populations $W_q(\mathbf{u}_1, \kappa_1), W_q(\mathbf{u}_2, \kappa_2)$ and $W_q(\mathbf{u}_3, \kappa_3)$, and we wish to test $H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 \pm \mathbf{u}_3 = \pm \mathbf{u}$, using the tabular, bootstrap and permutation

Table 8: Estimated significance levels (in %) of the tabular, bootstrap and permutation tests, for estimated concentration parameters, equal ($\kappa_1 = \kappa_2 = \kappa$) or not ($\kappa_1 \neq \kappa_2$) and several sample sizes n_1, n_2 .

n_1, n_2	κ	$q = 3$			$q = 4$		
		Tab.	Boot.	Perm.	Tab.	Boot.	Perm.
3, 5	1	4.4	8.1	5.6	5.6	8.8	6.0
	2	4.0	7.1	5.6	5.2	8.0	6.2
	5	2.8	4.6	5.3	3.7	5.3	5.7
	10	4.6	4.3	4.8	4.0	5.3	4.8
	20	4.0	5.5	4.8	4.3	6.2	4.8
5, 5	1	6.2	8.3	5.9	10.1	8.4	6.1
	2	4.7	7.7	5.9	8.5	7.6	6.2
	5	2.4	5.4	5.3	4.2	6.1	5.7
	10	3.5	5.1	5.0	4.2	6.0	5.1
	20	4.0	4.8	4.9	3.9	6.3	4.9
10, 10	1	12.4	8.1	6.6	20.0	7.4	6.6
	2	7.1	7.0	6.4	12.9	7.0	6.6
	5	2.5	6.6	5.3	3.1	6.4	5.8
	10	3.8	6.9	5.2	3.2	6.7	5.3
	20	4.6	5.7	5.2	4.4	6.7	5.2
n_1, n_2	κ_1, κ_2	$q = 3$			$q = 4$		
		Tab.	Boot.	Perm.	Tab.	Boot.	Perm.
3, 5	1,2	5.0	8.5	5.6	6.4	9.4	5.8
	3,5	3.4	5.2	5.3	5.9	5.8	5.7
	5,10	5.0	3.9	5.0	10.6	4.7	5.2
	10,20	5.8	4.1	4.9	8.3	4.7	4.9
5, 5	1,2	6.2	8.2	6.0	10.6	7.3	6.1
	3,5	3.8	5.6	5.5	7.0	5.8	6.0
	5,10	3.4	5.0	5.1	6.2	5.5	5.3
	10,20	4.2	4.4	4.9	3.9	5.8	5.1
10, 10	1,2	11.3	7.9	6.6	19.3	7.3	6.7
	3,5	3.7	6.0	5.5	6.5	6.1	5.9
	5,10	3.0	6.5	5.3	4.4	6.3	5.4
	10,20	4.4	6.0	5.2	4.3	6.8	5.2

Table 9: Empirical power (in %) of the tests for $p = 3, 4$, estimated concentrations κ_1, κ_2 ($\kappa_1 = \kappa_2 = \kappa$), angle θ ($^\circ$) and sample sizes n_1, n_2 .

n_1, n_2	p	θ/κ	Tabular				Bootstrap				Permutation			
			1	2	5	10	1	2	5	10	1	2	5	10
3, 5	3	18	7.6	7.5	7.0	16.6	45.6	37.6	18.2	19.2	5.6	5.6	5.3	4.9
		36	12.7	14.3	20.3	50.6	50.2	46.8	30.6	43.9	5.7	5.8	5.9	5.4
		54	18.9	24.6	41.5	83.4	51.8	52.7	45.9	63.4	5.6	5.9	6.6	5.1
		72	22.6	31.8	60.1	97.5	58.8	56.3	58.4	73.1	5.8	6.0	6.8	3.5
		90	25.2	35.8	64.7	99.1	55.8	58.4	64.5	77.2	6.1	5.9	6.8	3.0
	4	18	7.6	7.5	7.0	16.6	45.6	37.6	18.2	19.2	5.6	5.6	5.3	4.9
		36	12.7	14.3	20.3	50.6	50.2	46.8	30.6	43.9	5.7	5.8	5.9	5.4
		54	18.9	24.6	41.5	83.4	51.8	52.7	45.9	63.4	5.6	5.9	6.6	5.1
		72	22.6	31.8	60.1	97.5	58.8	56.3	58.4	73.1	5.8	6.0	6.8	3.5
		90	25.2	35.8	64.7	99.1	55.8	58.4	64.5	77.2	6.1	5.9	6.8	3.0
5, 5	3	18	6.5	5.5	6.4	25.7	38.4	30.7	16.8	26.9	5.9	5.9	5.7	5.1
		36	8.4	10.0	24.2	81.8	40.4	35.4	29.5	65.0	5.9	6.3	6.8	6.5
		54	12.7	20.2	56.0	99.1	44.3	43.8	50.9	84.3	6.0	6.8	9.3	11.5
		72	20.5	35.5	78.5	100.0	53.9	57.9	67.7	83.4	5.9	7.2	12.8	18.3
		90	27.2	45.0	85.8	100.0	58.6	62.9	69.7	77.6	5.7	7.3	14.5	19.9
	4	18	13.0	12.2	9.9	22.1	43.3	37.7	18.7	27.1	5.9	6.1	6.1	5.4
		36	20.2	22.8	31.1	68.4	48.6	45.4	37.4	60.5	5.8	6.2	7.6	6.9
		54	28.0	37.3	60.3	95.0	54.2	53.8	58.5	81.8	5.6	6.2	10.3	10.8
		72	34.0	47.6	79.5	99.9	58.4	61.1	74.4	89.6	5.3	6.4	13.5	17.2
		90	36.9	52.2	84.3	100.0	59.1	64.7	80.5	90.7	5.5	6.4	14.7	20.9
10, 10	3	18	13.9	9.9	13.4	55.0	31.8	22.3	19.3	44.4	6.5	6.5	5.6	5.4
		36	19.1	22.9	55.6	99.4	35.9	32.2	46.5	93.1	6.6	7.3	7.2	6.8
		54	30.2	47.7	91.4	100.0	46.0	49.8	78.5	99.4	6.9	8.6	10.4	10.6
		72	45.2	65.3	98.1	100.0	56.7	65.6	91.6	98.3	6.6	9.9	16.2	17.6
		90	51.1	75.0	98.8	100.0	62.6	72.7	93.3	96.2	6.6	9.4	19.9	31.8
	4	18	27.1	22.3	15.0	41.6	36.4	29.3	21.7	38.2	6.1	6.5	6.2	5.6
		36	41.5	46.7	61.3	95.6	44.1	42.3	55.3	87.5	5.5	6.5	8.2	7.2
		54	54.0	67.9	92.1	100.0	53.6	58.5	85.5	98.6	4.8	6.6	12.1	11.1
		72	62.4	79.4	98.0	100.0	58.0	65.8	96.5	99.7	4.3	6.5	17.7	19.8
		90	64.5	82.4	98.7	100.0	59.6	68.5	97.8	99.8	4.2	6.4	20.7	37.4

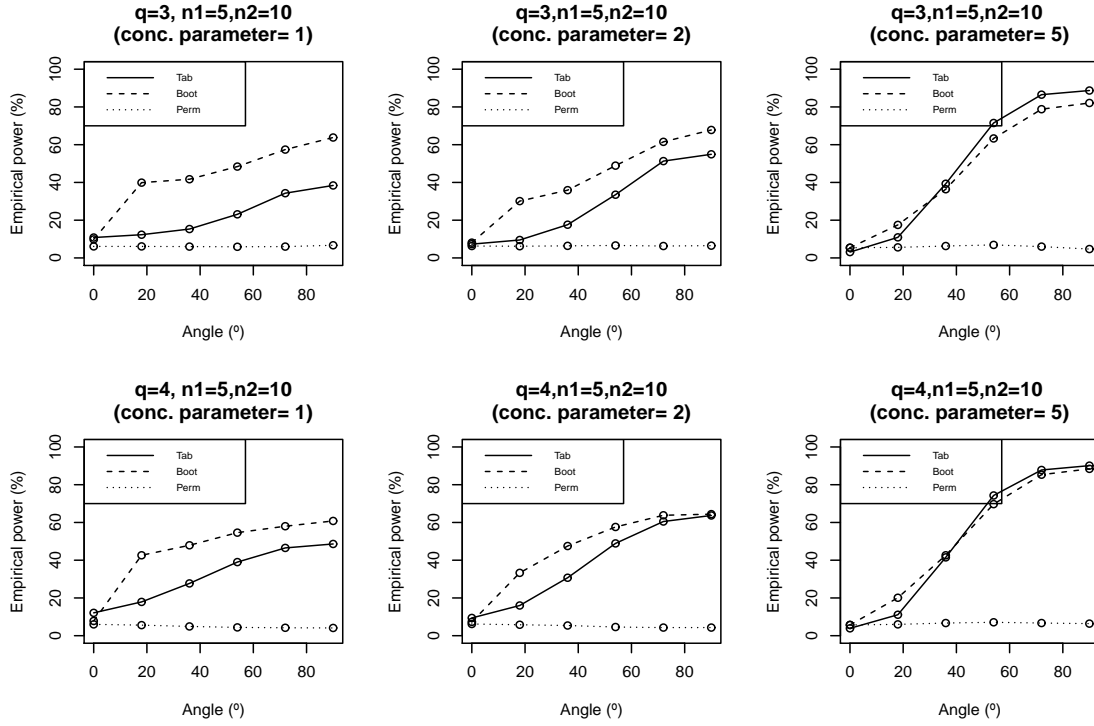


Figure 3: Empirical power of the tests for estimated equal concentrations

versions for the ANOVA test. We carried out a simulation study to estimate the level of significance and to determine the empirical power of the tests, considering the dimensions of the sphere $q = 3, 4$ and a common and known concentration parameter for the populations $\kappa_1 = \kappa_2 = \kappa_3 = \kappa = 1, 2, 5, 10$. We also considered equal samples size $n_1 = n_2 = n_3 = n = 5, 10$. We supposed, without loss of generality, that under $H_0 : \pm \mathbf{u}_1 = \pm \mathbf{u}_2 = \pm \mathbf{u}_3 = \pm \mathbf{e}_q$, where $\mathbf{e}_q = (0, \dots, 0, 1)'$. The estimated levels of significance were obtained for a nominal level of significance of 5% under H_0 . We determined the empirical power of the tests, for this nominal level of significance, supposing three types of alternative hypothesis. Let θ_1 be the angle between \mathbf{u}_1 and \mathbf{u}_2 , θ_2 be the angle between \mathbf{u}_2 and \mathbf{u}_3 and θ_3 has the same definition as θ_2 . We supposed, without loss of generality, in the alternative hypothesis: $H_1^{(1)}: \mathbf{u}_1 = \mathbf{e}_p, \mathbf{u}_2 = (0, \dots, 0, (1 - 0.95^2)^{1/2}, 0.95), \mathbf{u}_3 = (0, \dots, 0, (1 - 0.59^2)^{1/2}, 0.59)$, i.e, $\theta_1 = 18^\circ, \theta_2 = 54^\circ, \theta_3 = 36^\circ$, $H_1^{(2)}: \mathbf{u}_1 = \mathbf{e}_q, \mathbf{u}_2 = (0, \dots, 0, (1 - 0.95^2)^{1/2}, 0.95), \mathbf{u}_3 = \mathbf{e}_1$, i.e, $\theta_1 = 18^\circ, \theta_2 = \theta_3 = 90^\circ$ and $H_1^{(3)}: \mathbf{u}_1 = \mathbf{e}_q, \mathbf{u}_2 = \mathbf{e}_{q-1}, \mathbf{u}_3 = \mathbf{e}_1$, i.e, $\theta_1 = \theta_2 = \theta_3 = 90^\circ$. The number of replicates in the tests and the number of bootstrap or permutation samples considered to determine the levels of significance and the empirical power were the same as in the previous simulation study done for two populations. The estimated level of significance, obtained when $\theta_1 = \theta_2 = \theta_3 = 0^\circ$ and the empirical power for the three types of alternative hypothesis are indicated in Table 11. In this table we highlight the values of the power, in

Table 10: Empirical power (in %) of the tests, for $q = 3, 4$, estimated concentrations κ_1, κ_2 ($\kappa_1 \neq \kappa_2$), angle $\theta(^{\circ})$ and several sample sizes n_1, n_2 .

n_1, n_2	q	$\theta \backslash \kappa_1, \kappa_2$	Tabular			Bootstrap			Permutation		
			1,2	3,5	5,10	1,2	3,5	5,10	1,2	3,5	5,10
3, 5	3	18	6.3	4.9	11.1	41.3	19.6	18.6	5.7	5.3	5.1
		36	8.6	11.6	33.4	42.2	26.5	32.8	5.6	5.6	5.1
		54	12.6	24.1	63.0	46.8	39.1	45.4	5.7	5.6	4.3
		72	18.9	40.9	79.3	53.3	48.4	52.4	5.4	5.0	2.7
		90	22.8	47.8	84.0	56.5	49.5	50.1	5.3	4.1	2.0
	4	18	8.4	8.3	16.3	48.2	28.7	20.2	5.6	5.9	5.4
		36	13.6	18.9	32.5	52.5	37.6	33.6	5.4	6.0	5.7
		54	20.5	32.5	54.2	56.0	47.1	46.8	5.5	6.0	6.0
		72	24.8	47.5	72.2	61.7	58.0	57.9	5.4	5.9	5.6
		90	27.5	51.0	77.1	60.8	58.8	63.1	5.2	5.6	5.0
5, 5	3	18	7.1	6.6	12.3	38.5	20.5	20.4	5.9	5.6	5.4
		36	10.2	17.2	42.2	41.4	30.4	41.9	6.1	6.7	6.4
		54	15.4	35.5	71.0	46.2	43.3	59.7	6.2	7.9	8.8
		72	23.7	56.7	84.2	55.7	56.2	67.9	6.1	9.5	11.4
		90	28.8	61.9	87.4	58.8	60.4	66.2	6.0	9.8	12.9
	4	18	13.7	10.5	12.7	42.6	25.4	21.5	6.0	6.3	5.8
		36	20.7	24.8	36.2	46.8	34.4	41.1	5.9	7.1	7.0
		54	29.0	46.8	66.7	54.4	48.9	61.9	5.8	8.5	9.9
		72	35.9	63.7	84.8	57.6	59.7	77.1	5.6	9.5	13.2
		90	38.7	68.2	89.3	59.2	64.4	81.2	5.4	9.9	15.0
10, 10	3	18	14.0	10.7	23.5	30.5	18.7	24.8	6.6	5.9	5.6
		36	21.3	37.7	76.3	36.9	37.7	62.7	6.7	6.9	6.8
		54	34.5	70.8	95.8	48.0	59.9	87.1	7.0	9.1	9.0
		72	47.3	85.2	98.7	58.3	76.4	92.3	7.1	11.2	12.2
		90	52.3	87.7	98.7	63.1	80.0	90.9	6.9	11.0	14.1
	4	18	26.1	15.5	19.7	33.9	20.1	24.9	6.3	6.2	5.9
		36	40.9	48.0	66.1	41.3	40.1	60.1	5.7	7.7	7.4
		54	54.5	78.9	93.8	49.9	62.5	88.3	5.2	9.7	10.3
		72	64.2	90.9	98.4	56.5	78.6	97.0	4.8	11.3	13.9
		90	66.5	93.1	99.1	59.0	82.9	98.2	4.5	11.4	15.6

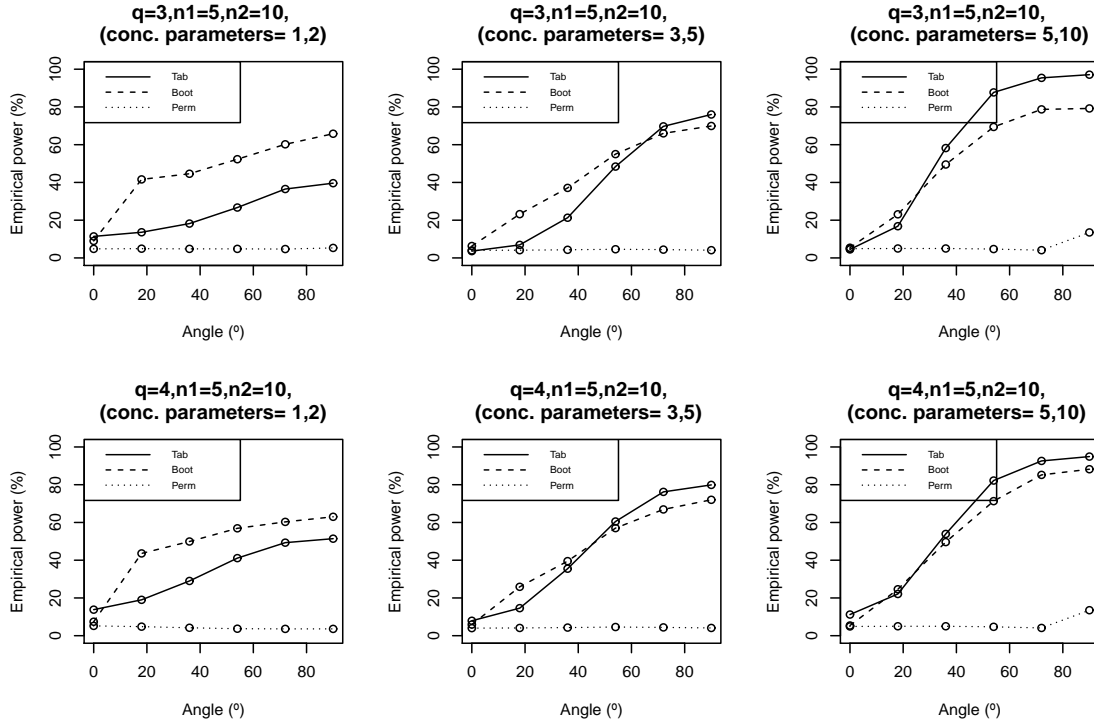


Figure 4: Empirical power of the tests for estimated different concentrations

which the bootstrap test is the most powerful.

The conclusions are similar to those obtained for two Watson populations, despite of the estimated levels of significance seem to be a bit worse. The estimated levels of significance in the bootstrap test are similar to the values for the permutation test, although in this latter test they are slightly better. In what concerns to the estimated level of significance for the tabular test, as this test is valid only for large concentrations, it would be expected that the estimated level of significance is not good for small concentrations.

Similarly, for each dimension of the sphere, the empirical power increases in general, as the separation between populations increases or the common concentration parameter increases.

We concluded that the bootstrap test is a good alternative to the tabular test for small concentration parameter or small samples or poor separation between the Watson populations. Among the three tests, the permutation test is the one that is least powerful, although it is the test that has in general the best estimated level of significance.

Table 11: Estimated significance level and empirical power (in %) of the tabular, bootstrap and permutation tests, for three Watson populations, with $q = 3$ and $q = 4$, common and known concentration parameter κ , common samples size n and angles between directional parameters θ_1 , θ_2 and θ_3 (in $^\circ$) .

Test	$q = 3$								$q = 4$			
	n	θ_1	θ_2	$\theta_3 \backslash \kappa$	1	2	5	10	1	2	5	10
Tabular		0	0	0	11.5	7.8	2.7	3.8	13.3	9.6	3.2	2.9
	5	18	54	36	18.9	23.4	51.2	99.1	26.6	31.7	46.5	91.3
		18	90	90	18.8	33.1	81.5	100.0	20.2	30.4	70.5	100.0
		90	90	90	21.5	42.2	93.7	100.0	42.3	60.5	92.7	100.0
		0	0	0	23.8	11.1	2.3	3.2	31.9	16.6	2.6	2.8
	10	18	54	36	44.5	52.4	90.6	100.0	58.2	68.0	89.0	99.9
		18	90	90	45.8	70.6	99.2	100.0	52.1	69.7	97.3	100.0
		90	90	90	53.1	84.1	100.0	100.0	80.2	93.8	100.0	100.0
Bootstrap		0	0	0	7.1	6.6	5.6	5.7	7.5	6.9	5.7	5.6
	5	18	54	36	62.0	57.7	59.6	94.6	70.8	67.1	59.5	82.2
		18	90	90	60.6	61.5	78.6	99.7	66.8	64.9	72.2	99.4
		90	90	90	62.8	67.4	88.2	100.0	66.4	67.9	81.9	100.0
		0	0	0	6.9	6.2	6.1	6.7	7.3	6.4	6.1	6.5
	10	18	54	36	49.2	49.9	79.4	99.9	58.7	58.8	73.9	97.8
		18	90	90	49.2	59.9	94.7	100.0	52.9	57.7	88.2	100.0
		90	90	90	53.5	70.5	98.7	100.0	51.4	63.3	95.6	100.0
Permutation		0	0	0	6.2	6.4	5.6	5.2	6.2	6.5	6.1	5.3
	5	18	54	36	6.5	7.6	9.2	8.3	5.2	6.4	10.1	8.2
		18	90	90	7.2	10.3	19.5	11.5	6.4	8.8	19.3	15.3
		90	90	90	7.3	11.3	27.7	27.6	6.7	9.5	26.5	36.1
		0	0	0	7.0	6.9	5.4	5.3	6.8	7.1	5.8	5.3
	10	18	54	36	7.6	9.3	10.2	8.8	4.2	6.5	11.3	8.7
		18	90	90	9.6	18.0	28.2	13.1	7.8	14.8	30.8	15.7
		90	90	90	10.6	23.5	54.8	56.8	8.6	19.0	54.5	64.0

5 Application

We used the vectorcardiogram data of Downs *et al.* (1971) obtained with two systems (Frank system and McFee lead system). From these data we took the unit spherical vector

associated with each vectorcardiogram, which represents the spatial direction of the vector of the QRS loop having the greatest magnitude. Then, we considered the axes associated to these directions. We selected data for eight children from each of the eight combinations of the categories (sex-age and type of system). Data, in radians, are in the Table 12.

Table 12: Spherical vectorcardiogram data (in radians)

		Frank system				McFee lead system			
Boy aged 2-10		0.587	0.706	0.614	0.744	-0.052	0.584	0.597	0.873
		0.645	0.568	0.614	0.547	0.130	0.809	0.609	0.475
		0.489	0.422	0.495	0.383	0.990	0.067	0.522	-0.110
		0.552	0.863	0.145	0.160	0.959	0.553	0.678	0.927
		0.702	0.410	0.814	0.261	0.246	0.651	0.713	0.273
		0.449	0.296	0.562	0.952	0.143	0.520	0.178	0.259
Boy aged 11-19		0.536	0.561	-0.167	0.265	0.718	0.770	0.755	0.712
		0.572	0.662	0.481	0.805	0.694	0.511	0.358	0.702
		0.621	0.497	0.861	0.531	-0.049	0.382	0.550	0.031
		0.470	0.101	0.256	0.658	0.623	0.809	0.903	0.509
		0.690	0.404	0.735	0.596	0.543	0.586	0.267	0.581
		0.550	0.909	0.628	0.460	0.563	0.035	-0.337	-0.636
Girl aged 2-10		0.404	0.526	0.721	0.581	0.857	0.255	0.882	0.781
		0.616	0.570	0.573	0.752	0.375	0.964	0.070	0.624
		0.676	0.631	0.389	0.311	0.354	0.073	0.466	0.027
		0.390	0.595	0.560	0.536	0.136	0.941	0.694	0.769
		0.711	0.618	0.797	0.515	0.034	0.339	0.694	0.481
		0.585	0.515	0.226	0.670	-0.990	-0.025	0.188	0.421
Girl aged 11-19		0.472	0.453	0.701	0.286	0.778	0.674	0.722	0.583
		0.489	0.717	0.574	0.622	0.589	0.628	0.578	0.793
		0.734	0.530	0.423	0.729	0.217	-0.388	0.380	0.176
		0.108	0.299	0.535	0.395	0.787	0.680	0.693	0.525
		0.851	0.713	0.720	0.642	0.537	0.580	0.707	0.766
		0.513	0.634	0.442	0.657	0.303	0.448	0.140	0.372

We are interested in investigating whether for each sex-age category, the type of system (Frank system or McFee lead system) affects the result of the vectorcardiogram. In this application we supposed the ANOVA statistic for different concentration parameters (general statistic) and also the ANOVA statistic for equal concentration parameters for each sex-age category. Then for each category sex-age, we determined the values of the ANOVA

Table 13: Largest eigenvalues and estimates of the concentration parameters of the groups, and statistic values and p-values of the tests for each sex-age category

Sex-Age	Group	System		Concen- tration parameters	Statistic value	p-value (%)		
		Frank	McFee lead			Tab.	Boot.	Perm.
	j	1	2					
Boy aged 2-10	\hat{w}_j	7.087	6.302	Different	1.266	29.8	51.7	19.8
	$\hat{\kappa}_j$	9.454	5.500	Equal	1.297	28.9	51.3	20.7
Boy aged 11-19	\hat{w}_j	7.173	6.561	Different	9.316	0.1	0.1	0.1
	$\hat{\kappa}_j$	10.343	6.360	Equal	9.538	0.1	0.1	0.1
Girl aged 2-10	\hat{w}_j	7.648	6.013	Different	2.441	10.5	34.6	3.4
	$\hat{\kappa}_j$	23.251	4.760	Equal	2.673	8.6	39.5	2.8
Girl aged 11-19	\hat{w}_j	7.589	7.419	Different	11.244	0	0.1	0.1
	$\hat{\kappa}_j$	20.030	14.360	Equal	11.299	0	0.1	0.1

statistics given by (2.10) and (2.12), which are indicated in Table 13, as well as the p-values obtained for the tabular method, the bootstrap and permutation versions of the ANOVA 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 children aged 11-19 and boys aged 2-10. More precisely, we can conclude that there is no significant difference between the systems for boys aged 2-10 while there is difference for children aged 11-19. On the other hand, for girls aged 2-10 there is no evidence to conclude that the systems differ using the tabular and bootstrap tests. Based on the permutation test, we can not conclude that the systems differ at a level of significance 1%, but we conclude that there is difference between the systems at a level of 5%. The code for applying these tests is available in the web page: https://sigarra.up.pt/fep/pt/conteudos_geral.ver?pct_pag_id=1010326&pct_parametros=p_codigo=205276&pct_grupo=23660#23660

6 Concluding remarks

We have concluded that the bootstrap and permutation versions of the ANOVA 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 three tests, the bootstrap test is in general the most powerful test in the case of small samples for small concentrations or bad separation between the Watson populations. So, in these cases, the bootstrap and permutation tests based on ANOVA statistic may constitute useful alternatives to the ANOVA statistic, that has an asymptotic distribution, valid only for large concentrations.

Acknowledgements

- The author is grateful to the Professors Michael Stephens and Richard Lockhart for their suggestions to this paper. The author also thanks the helpful comments given by the referees of this journal, that helped to improve this paper.
- This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

- [1] 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*, 102:478, 695-707.
- [2] 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.
- [3] Downs, T., Liebman, J. and Mackay, W. (1971). Statistical methods for vectorcardiogram orientations. In: Hoffman, R. I., Glassman, E. (eds). *Vectorcardiography 2: proceedings of XI th international symposium on vectorcardiography*. North-Holland, Amsterdam, 216-222.
- [4] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1, 1-26.
- [5] Fisher, N. I. (1993). *Statistical analysis of circular data*, Cambridge University Press, Cambridge, Great Britain.

- [6] Fisher, N. I. and Hall, P. (1989). Bootstrap Confidence Regions for Directional Data. *Journal of the American Statistical Association*, 84:408, 996-1002.
- [7] Fisher, N. I., Hall, P., Jing, B.-Y. and Wood, A. T. A. (1996). Improved Pivotal Methods for Constructing Confidence Regions with Directional Data. *Journal of the American Statistical Association*, 91, 1062-1070.
- [8] Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987). *Statistical analysis of spherical data*, Cambridge University Press, Cambridge, Great Britain.
- [9] Gomes, P. and Figueiredo, A. (1999). A new probabilistic approach for the classification of normalised variables, *Bulletin of the International Statistical Institute*, vol. LVIII, n^o 1, p. 403-404.
- [10] Good, P. (2004). *Permutation, Parametric and Bootstrap Tests of Hypotheses*, New York: Springer-Verlag.
- [11] Harrison, D., Kanji, G. K. and Gadsden, R. J. (1986). Analysis of variance for circular data. *Journal of Applied Statistics*, 13, 123-138.
- [12] Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific: Singapore.
- [13] Li, K.- H. and Wong, C. K. - F. (1993). Random sampling from the Watson distribution. *Communications in Statistics - Computation and Simulation*, 22, (4), 997-1009.
- [14] Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. John Wiley and Sons, Chichester.
- [15] Romano, J. P. (1990). On behavior of randomization tests without the group invariance assumption. *Journal of the American Statistical Association*, 85, 686-692.
- [16] Stephens, M. A. (1969). Multi-sample tests for the Fisher distribution for directions, *Biometrika*, 56, 1, 169-181.
- [17] Stephens, M. A. (1992). On Watson's ANOVA for directions. In Watson, G. and Mardia K. V. (eds.) *Art of Statistical Science*, 75-85, Wiley, University of Michigan.
- [18] Underwood, A. J. and Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology*, 91, 17-43.
- [19] Watson, G. S. (1983). *Statistics on spheres*. John Wiley and Sons, New York.
- [20] Wellner, J. A. (1979). Permutation tests for directional data. *The Annals of Statistics*, 7, 929-943.