Discriminant Analysis for the von Mises-Fisher distribution

Adelaide Maria de Sousa Figueiredo

Discriminant Analysis for the von Mises-Fisher distribution

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2008-0319.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	11-Jul-2009
Complete List of Authors:	Figueiredo, Adelaide; Faculdade de Economia da Universidade do Porto, Grupo de Matemática e Informática
Keywords:	Directional data, Discriminant Analysis, von Mises-Fisher distribution
Abstract:	The von Mises-Fisher distribution is widely used for modeling directional data. In this paper we derive the discriminant rules based on this distribution to assign objects into pre-existing classes. We determine a distance between two von Mises-Fisher populations and we calculate estimates of the misclassification probabilities. We also analyse the behavior of the distance between two von Mises-Fisher populations and of the estimates of the misclassification probabilities when we modify the parameters of the populations or the samples size or the dimension of the sphere. Finally, we present an example with real spherical data available in the literature.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

DiscriminantVMisesAdelaideFigueiredo.zip

Discriminant Analysis for the von Mises-Fisher distribution

Adelaide Figueiredo

Faculdade de Economia / LIAAD - INESC Porto L.A., Universidade do Porto

Rua Dr. Roberto Frias, 4200-464 Porto - PORTUGAL

Abstract. The von Mises-Fisher distribution is widely used for modeling directional data. In this paper we derive the discriminant rules based on this distribution to assign objects into pre-existing classes. We determine a distance between two von Mises-Fisher populations and we calculate estimates of the misclassification probabilities. We also analyse the behavior of the distance between two von Mises-Fisher populations and of the estimates of the misclassification probabilities when we modify the parameters of the populations or the samples size or the dimension of the sphere. Finally, we present an example with real spherical data available in the literature.

Keywords. *Directional data; Discriminant Analysis; von Mises-Fisher distribution.*

AMS 2000 subject classification. 62H11, 62H30.

1 Introduction

Directional statistics has been developed by many authors, for instance by Watson (1983), Mardia and Jupp (2000), Fisher, Lewis and Embleton (1987), Fisher (1993), Stephens (1982, 1992) and Jammalamadaka and SenGupta (2001). Discriminant Analysis for directional data was treated in the literature by Morris and Laycock (1974) and by El Khattabbi and Streit (1996) for circular data and spherical data. Figueiredo and Gomes (2006) considered discriminant analysis based on the Watson distribution defined on the hypersphere. The aim of this paper is to give discriminant rules based on the von Mises-Fisher distribution. We also determine a distance between two von Mises-Fisher populations with the same concentration parameter or distinct concentration parameters. As the misclassification probabilities cannot be obtained in closed form in the case of von Mises-Fisher populations, we determine estimates of these probabilities for several dimensions of the sphere, by generating samples from the von Mises-Fisher populations. For the particular case of the Fisher distribution, Morris and Laycock (1974) determined some expressions for the misclassification probabilities. We also analyse the behavior of the distance between the populations and of the misclassification probabilities estimates when we modify the parameters of the populations or the samples size or the dimension of the sphere.

In Sec. 2 we recall the von Mises-Fisher distribution; in Sec. 3 we consider discriminant analysis for the von Mises-Fisher distribution; in Sec. 4 we determine the distance between two von Mises-Fisher populations and the estimates of the misclassification probabilities for several values of the parameters of the populations, several samples sizes and several dimensions of the spheres; and finally in Sec. 5 we give an example with real spherical data.

2 von Mises-Fisher distribution

The von Mises-Fisher distribution on the unit sphere in $\mathbb{R}^p$, $S_{p-1} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}'\mathbf{x} = 1\}$ is usually denoted by $M_p(\boldsymbol{\mu}, \kappa)$ and has probability density function given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_p(\kappa) \exp\left(\kappa \mathbf{x}'\boldsymbol{\mu}\right) \quad \mathbf{x} \in S_{p-1}, \quad \boldsymbol{\mu} \in S_{p-1}, \ \kappa \geq 0\ , \tag{2.1}$$

where the normalising constant is given by

$$c_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \tag{2.2}$$

with I_ν denoting the modified Bessel function of the first kind and order ν defined by

$$I_\nu(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos \nu\theta e^{\kappa \cos\theta} \mathrm{d}\theta. \tag{2.3}$$

The parameter $\boldsymbol{\mu}$ is the mean direction and the parameter κ specifies the concentration around $\boldsymbol{\mu}$. The parameter $\boldsymbol{\mu}$ is also the mode and $-\boldsymbol{\mu}$ is the antimode (provided that $\kappa > 0$). The distribution is rotationally symmetric about $\boldsymbol{\mu}$ (Mardia and Jupp, 2000, p. 179).

The distribution $M_p(\boldsymbol{\mu}, \kappa)$ for the particular case of $p = 3$ is called the Fisher distribution and for the particular case of $p = 2$ is called the von Mises distribution.

If $\mathbf{x}$ comes from $M_p(\boldsymbol{\mu}, \kappa)$ and U is a non-random orthogonal matrix, then $U\mathbf{x}$ comes from $M_p(U\boldsymbol{\mu}, \kappa)$.

If $\mathbf{x}$ comes from $M_p(\boldsymbol{\mu}, \kappa)$ then for large κ:

$$2\kappa\left(1 - \mathbf{x}'\boldsymbol{\mu}\right) \dot{\sim} \chi^2_{p-1}. \tag{2.4}$$

(See Mardia and Jupp, 2000, p. 172).

Let $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ be a random sample of size n from the von Mises-Fisher distribution $M_p(\boldsymbol{\mu}, \kappa)$. Let $\overline{R}$ be the mean resultant length of the sample defined by $\overline{R} = \|\overline{\mathbf{x}}\| = (\overline{\mathbf{x}}'\overline{\mathbf{x}})^{1/2}$,

where $\overline{\mathbf{x}}$ is the sample vector mean of $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ defined by $\overline{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i / n$.

The maximum likelihood estimator of $\boldsymbol{\mu}$ is the sample mean direction, that is $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}} / \|\overline{\mathbf{x}}\|$ and the maximum likelihood estimator of κ is the solution of the equation $A_p(\kappa) = \|\overline{\mathbf{x}}\|$, where the function $A_p(\kappa)$ is defined by

$$A_p(\kappa) = \frac{c_p'(\kappa)}{c_p(\kappa)} = \frac{I_{\frac{p}{2}}(\kappa)}{I_{\frac{p}{2}-1}(\kappa)}. \tag{2.5}$$

(See Mardia and Jupp, 2000, p. 198).

3 Discriminant Analysis

We suppose without loss of generality, two von Mises-Fisher populations, $P_1 : M_p(\boldsymbol{\mu}_1, \kappa_1)$ and $P_2 : M_p(\boldsymbol{\mu}_2, \kappa_2)$. Let $(\mathbf{x}_{i1}, ..., \mathbf{x}_{in_i})$ be the sample from the population i, $i = 1, 2$. Next, we suppose that the parameters of the populations $(\boldsymbol{\mu}_i, \kappa_i)$, $i = 1, 2$ are known. When these parameters are unknown, it is usual to replace them by their maximum likelihood estimates obtained from the respective samples.

The density function of the ith population is given by

$$f_i(\mathbf{x}) = c_p(\kappa_i) \exp\left(\kappa_i \boldsymbol{\mu}_i' \mathbf{x}\right), \quad \mathbf{x} \in S_{p-1},\ \kappa_i \geq 0,\ \boldsymbol{\mu}_i \in S_{p-1}, \tag{3.1}$$

where the normalising constant $c_p(\kappa_i)$ is defined by (2.2).

Our aim is to define a discriminant rule to assign a new object into one of the populations, that is, we want to divide the discriminant space into mutually exclusive and exhaustive regions R_1 and R_2, in order that any given object is classified into the population into whose region it falls.

Let π_i $(i = 1, 2)$ be the *prior* probability associated with the ith population; $C(i|j)$ is the cost of classifying incorrectly an object from the jth population into the ith population $(i,$

$j = 1, 2,\ i \neq j$); $P\left(i|j\right)$ is the probability of misclassifying an object from the jth population into the ith population ($i,\ j = 1, 2,\ i \neq j$) defined by

$$P\left(i|j\right) = P\left(\mathbf{x} \in R_i|\mathbf{x} \in P_j\right) = \int_{R_i} f_j\left(\mathbf{x}\right) \mathrm{d}\mathbf{x}; \tag{3.2}$$

and C be the total cost of misclassification.

The Bayes rule that minimises the expected cost of misclassification $E(C)$ defined by $E(C) = C(2|1)P(2|1)\pi_1 + C(1|2)P(1|2)\pi_2$ is the following:

Assign $\mathbf{x}$ to population 1 if

$$C(2|1) f_1\left(\mathbf{x}\right) \pi_1 \geq C(1|2) f_2\left(\mathbf{x}\right) \pi_2 \tag{3.3}$$

and to population 2, otherwise (see for instance, Johnson and Wichern, 1992, p. 557).

Substituting for $f_1\left(\mathbf{x}\right)$ and $f_2\left(\mathbf{x}\right)$ in (3.3), the Bayes rule is given by

Assign $\mathbf{x}$ to $M_p\left(\boldsymbol{\mu}_1, \kappa_1\right)$ if

$$\left(\kappa_1\boldsymbol{\mu}_1 - \kappa_2\boldsymbol{\mu}_2\right)' \mathbf{x} + \ln \frac{c_p\left(\kappa_1\right)}{c_p\left(\kappa_2\right)} \geq \ln \frac{C\left(1|2\right)\pi_2}{C\left(2|1\right)\pi_1} \tag{3.4}$$

and to $M_p\left(\boldsymbol{\mu}_2, \kappa_2\right)$ otherwise.

Denoting the discriminant function by W_{12}, i.e.

$$W_{12} = \left(\kappa_1\boldsymbol{\mu}_1 - \kappa_2\boldsymbol{\mu}_2\right)' \mathbf{x} + \ln \frac{c_p\left(\kappa_1\right) C\left(2|1\right)\pi_1}{c_p\left(\kappa_2\right) C\left(1|2\right)\pi_2}, \tag{3.5}$$

the discriminant regions R_1 and R_2 are given as

$$\begin{aligned} R_1 &= \left\{\mathbf{x} \in S_{p-1} : W_{12} \geq 0\right\} \\ R_2 &= \left\{\mathbf{x} \in S_{p-1} : W_{12} < 0\right\}. \end{aligned}$$

A measure of the distance between two populations, referred in Morris and Laycock (1974) and used by El Khattabi and Streit (1996), which is equivalent to Mahalonobis's D^2

in the multivariate normal case, is defined by

$$D = E\left(W_{12}|\mathbf{x} \in P_1\right) - E\left(W_{12}|\mathbf{x} \in P_2\right). \tag{3.6}$$

If $\mathbf{x}$ comes from the von Mises-Fisher distribution $M_p\left(\boldsymbol{\mu}, \kappa\right)$, then

$$E\left(\mathbf{x}\right) = \rho\boldsymbol{\mu},$$

where $\rho = A_p\left(k\right)$ with the function $A_p\left(\kappa\right)$ defined by (2.5) (Mardia and Jupp, 2000, p. 169). Note that for $p = 3$, the function $A_p\left(k\right)$ may be calculated as $A_p\left(k\right) = \coth\kappa - \frac{1}{\kappa}$ (Mardia and Jupp, 2000, p. 169).

Then,

$$\begin{aligned} E\left(W_{12}|\mathbf{x} \in P_1\right) &= \ln\frac{c_p\left(\kappa_1\right)C\left(2|1\right)\pi_1}{c_p\left(\kappa_2\right)C\left(1|2\right)\pi_2} + A_p\left(\kappa_1\right)\left(\kappa_1\boldsymbol{\mu}_1 - \kappa_2\boldsymbol{\mu}_2\right)'\boldsymbol{\mu}_1 \\ E\left(W_{12}|\mathbf{x} \in P_2\right) &= \ln\frac{c_p\left(\kappa_1\right)C\left(2|1\right)\pi_1}{c_p\left(\kappa_2\right)C\left(1|2\right)\pi_2} + A_p\left(\kappa_2\right)\left(\kappa_1\boldsymbol{\mu}_1 - \kappa_2\boldsymbol{\mu}_2\right)'\boldsymbol{\mu}_2 \end{aligned}$$

and consequently, the distance D is given by

$$D = \left(\kappa_1\boldsymbol{\mu}_1 - \kappa_2\boldsymbol{\mu}_2\right)'\left(A_p\left(\kappa_1\right)\boldsymbol{\mu}_1 - A_p\left(\kappa_2\right)\boldsymbol{\mu}_2\right). \tag{3.7}$$

If the two populations have the same concentration parameter, i.e., $\kappa_1 = \kappa_2 = \kappa$, the distance D simplifies to

$$D = \kappa A_p\left(\kappa\right)\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)'\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right). \tag{3.8}$$

The probabilities of misclassification are defined by

$$P\left(1|2\right) = P\left(W_{12} \geq 0|\mathbf{x} \in P_2\right)$$

$$P\left(2|1\right) = P\left(W_{12} < 0|\mathbf{x} \in P_1\right).$$

As we can not obtain any closed expressions for these misclassification probabilities in the case of the von Mises-Fisher distribution, we estimate them in some cases in the next

section. These probabilities can also be represented, as an integral, which can be calculated numerically. Morris and Laycock (1974) expressed these misclassification probabilities in terms of an integral in the cases of $p = 2$ and $p = 3$. In the next section we determine the misclassification probabilities for $p = 2$. For the calculation of the integrals we have used the algorithm proposed in Hill (1977) or alternatively, we could have used the tables of the distribution function for the von Mises distribution given for instance, in Mardia and Jupp (2000), Appendix 2.

For k von Mises-Fisher populations, the discriminant functions are given by

$$W_{ji} = \left(\kappa_j \boldsymbol{\mu}_j - \kappa_i \boldsymbol{\mu}_i\right)' \mathbf{x} + \ln \frac{c_p\left(\kappa_j\right) C\left(i|j\right) \pi_j}{c_p\left(\kappa_i\right) C\left(j|i\right) \pi_i}, \quad i, j = 1, ..., k, \; i \neq j \tag{3.9}$$

and the discriminant regions are defined by

$$R_j = \left\{\mathbf{x} \in S_{p-1} : W_{ji} \geq 0, \; i = 1, ..., k, \; i \neq j\right\}.$$

It is enough to use $k - 1$ functions for defining the discriminant regions, provided that $k - 1 < p$; we may use for instance, the discriminant functions W_{1j}, $j = 2, ..., k$.

We have not presented the previous results for the particular cases of $p = 2$ and $p = 3$ because they were already considered in Morris and Laycock (1974).

4 Estimates of the misclassification probabilities and distance between the two populations

We have investigated the above procedures by a large Monte Carlo study. We have considered two von Mises-Fisher populations $P_1 : M_p\left(\boldsymbol{\mu}_1, \kappa_1\right)$ and $P_2 : M_p\left(\boldsymbol{\mu}_2, \kappa_2\right)$, with known parameters. We have supposed the two populations with equal *prior* probabilities: $\pi_1 = \pi_2 = 0.5$ and equal costs of misclassification: $C(1|2) = C(2|1)$. We have considered the

following dimensions of the sphere: $p = 2, 3, 4, 10,$ and for each p, we have taken the mean directions: $\boldsymbol{\mu}_1 = (0, ..., 0, 1)'$ and $\boldsymbol{\mu}_2 = \left(0, 0, ..., \left(1 - \cos^2\theta\right)^{1/2}, \cos\theta\right)'$, supposing that the angle θ between these directions is equal to $0^o(30^o)180^o$. The results only depend on $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ through their angular separation $\theta = acos(\boldsymbol{\mu}_1'.\boldsymbol{\mu}_2)$. We have supposed the following two cases:

- Populations with equal concentration parameters: $\kappa_1 = \kappa_2 = \kappa$, with $\kappa = (3), 5, 10, 15(20)$.
- Populations with different concentration parameters: $\kappa_1 \neq \kappa_2$, $\kappa_1 = 5, \kappa_2 = 10, 15, 20(25)$. (We have considered only $\kappa_1 < \kappa_2$, because if $\kappa_1 > \kappa_2$, the distance D and the estimates of the misclassification probabilities are identical).

For both cases, we have estimated the misclassification probabilities $P(2|1)$ and $P(1|2)$ by the misclassification proportions in 10 000 observations generated from P_1 and P_2 populations, respectively, and then we have calculated the average of these proportions to obtain an estimate of the total misclassification probability. When the two populations have the same concentration parameter $\kappa_1 = \kappa_2$, then the misclassification probabilities $P(2|1)$ and $P(1|2)$ are equal (or approximately equal, if these are estimated), but when $\kappa_1 \neq \kappa_2$, the misclassification probabilities are not equal because in this case, the probability of assigning an observation to the population with greater concentration parameter is the larger (see the expression (4.2)).

For generating observations from the von Mises-Fisher distribution, we have used the method given in Wood (1994).

4.1 Populations with equal concentration parameters

In this case the discriminant function defined by (3.5) reduces to

$$W_{12} = \kappa \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)' \mathbf{x}. \tag{4.1}$$

The values obtained for the distance D defined by (3.8) and for the estimates of the misclassification probabilities for some values of the angle $\theta(^o) = 30, 60, (90)120, (180)$ and $\kappa = (3), 5, 10, 15, (20)$ are indicated for $p = 2, 3, 4$ in Tables 1-3. See also figure 1.

As expected, for each p and fixed θ, when κ increases, the estimated misclassification probabilities decrease, tending very quickly to 0 and the distance D between the populations increases.

For each p and fixed κ when the angle θ increases, the estimates of the misclassification probabilities decrease and the distance D between the populations increases.

For fixed κ and θ, as the dimension of the sphere p increases, the estimated misclassification probabilities increase and the distance D between the populations decreases.

We have also observed that in the case of $p = 2$, for each θ, κ, the estimated misclassification probabilities are close from the theorical misclassification probabilities.

4.2 Populations with different concentration parameters

In this case the discriminant function defined by (3.5) reduces to

$$W_{12} = \left(\kappa_1 \boldsymbol{\mu}_1 - \kappa_2 \boldsymbol{\mu}_2\right)' \mathbf{x} + \ln \frac{c_p\left(\kappa_1\right)}{c_p\left(\kappa_2\right)}. \tag{4.2}$$

The values obtained for the distance D defined by (3.7) and for the estimates of the misclassification probabilities for for some values of the angle $\theta(^o) = 30, 60, (90)120, (180)$ and $\kappa_1 = 5, \kappa_2 = 10, 15, 20(25)$, are indicated for $p = 2, 3, 4$ in Tables 4-6. See also figure 2.

Table 1: Misclassification probabilities and distance D between the two populations for $p = 2$

$\theta(^o)$	κ	$\widehat{P(1\|2)} = \widehat{P(2\|1)}$	$P(1\|2) = P(2\|1)$	D
30	3	0.384	0.334	0.036
	5	0.278	0.285	0.317
	10	0.156	0.208	1.309
	15	0.099	0.158	2.374
60	3	0.187	0.197	1.299
	5	0.094	0.131	3.017
	10	0.026	0.053	7.431
	15	0.009	0.023	11.947
90	3	0.098	0.104	3.556
	5	0.032	0.049	7.192
	10	0.003	0.009	16.480
	15	0	0.002	25.914
120	3	0.056	0.049	6.392
	5	0.012	0.015	12.252
	10	0	0.001	26.722
	15	0	0	41.263

For each dimension of the sphere p, for fixed κ_1 and θ, when κ_2 increases, the misclassification probabilities estimates decrease and the distance D between the populations

Table 2: Estimates of the misclassification probabilities and distance D between the two populations for $p = 3$

$\theta(^o)$	κ	$\widehat{P(1\|2)}$	$\widehat{P(2\|1)}$	Average	D
	5	0.298	0.298	0.298	1.072
30	10	0.216	0.215	0.216	2.411
	15	0.162	0.161	0.162	3.751
	20	0.126	0.126	0.126	5.091
	5	0.149	0.145	0.147	4.000
60	10	0.060	0.059	0.060	9.000
	15	0.026	0.026	0.026	13.999
	20	0.012	0.012	0.012	18.999
	5	0.024	0.023	0.024	12.001
120	10	0.001	0.001	0.001	27.000
	15	0	0	0	41.999
	20	0	0	0	56.999
	5	0.007	0.007	0.007	16.002
180	10	0	0	0	36.000
	15	0	0	0	56.000
	20	0	0	0	76.000

increases.

For each p, for fixed κ_1 and κ_2, as the angle θ increases, the estimated misclassification probabilities decrease and the distance D between the populations increases.

Table 3: Estimates of the misclassification probabilities and distance D between the two populations for $p = 4$

$\theta(^o)$	κ	$\widehat{P(1\|2)}$	$\widehat{P(2\|1)}$	Average	D
	5	0.308	0.308	0.308	0.963
30	10	0.219	0.223	0.221	2.288
	15	0.165	0.167	0.166	3.625
	20	0.129	0.130	0.130	4.963
	5	0.158	0.160	0.159	3.595
60	10	0.062	0.065	0.064	8.549
	15	0.027	0.028	0.028	13.529
	20	0.013	0.014	0.014	18.519
	5	0.032	0.033	0.033	10.785
120	10	0.002	0.002	0.002	25.620
	15	0	0	0	40.589
	20	0	0	0	55.559
	5	0.012	0.012	0.012	14.380
180	10	0	0	0	34.160
	15	0	0	0	54.120
	20	0	0	0	74.080

For fixed $\kappa_1, \kappa_2, \theta$, when p increases, the estimates of the misclassification probabilities increase and the distance D between the populations decreases.

We have also observed that in the case of $p = 2$, for each $\theta, \kappa_1, \kappa_2$ and unless $\theta = 30^o$,

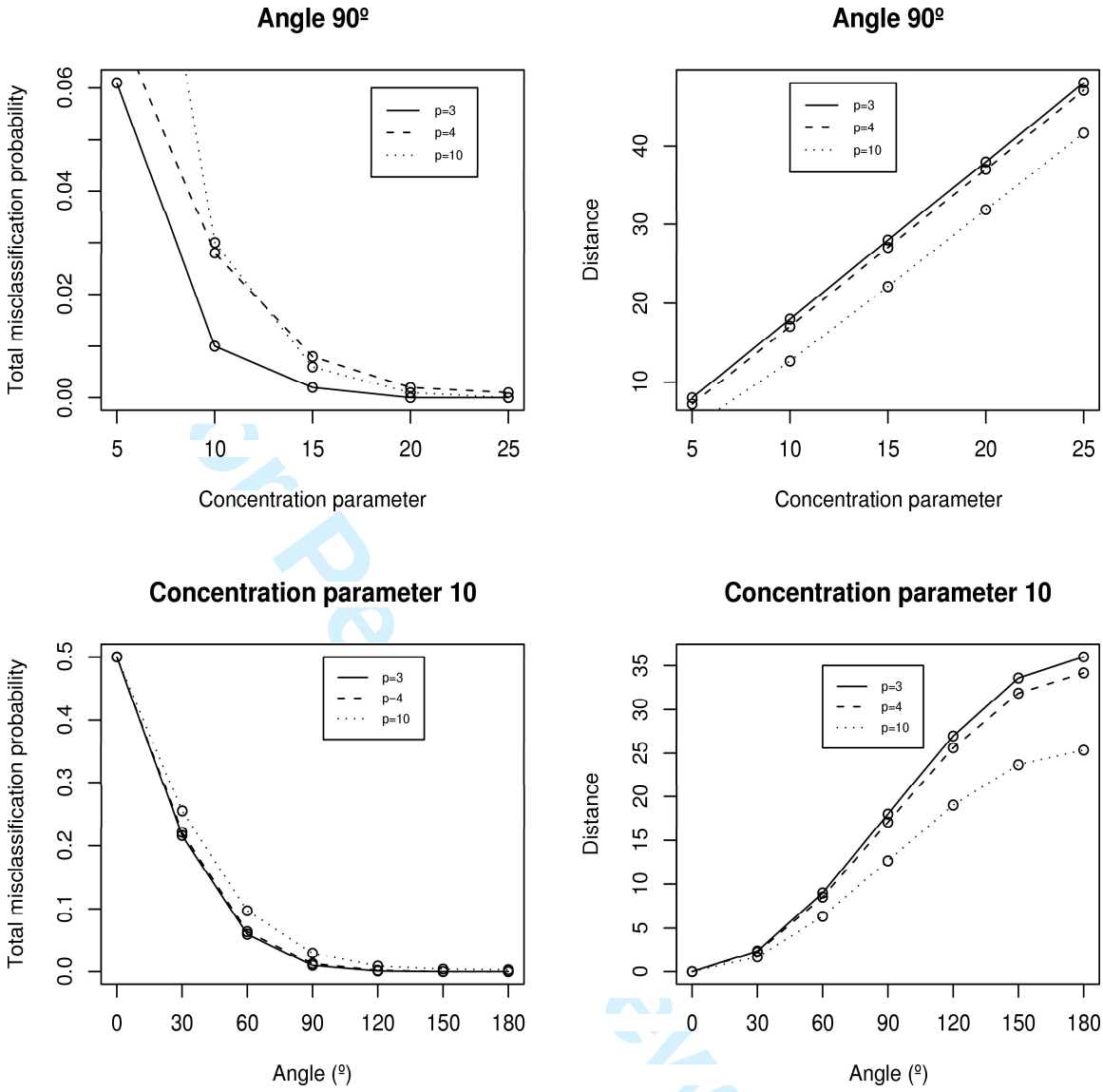

Figure 1: Estimate of the total misclassification probability and distance D between the two populations

the estimated misclassification probabilities are close from the theorical misclassification probabilities.

Table 4: Misclassification probabilities and distance D between the two populations for $p = 2$

$\theta(^o)$	κ_1	κ_2	$\widehat{P(1\|2)}$	$\widehat{P(2\|1)}$	Average	$P(1\|2)$	$P(2\|1)$	Average	D
		10	0.114	0.144	0.129	0.253	0.321	0.287	1.141
30	5	15	0.069	0.084	0.077	0.200	0.312	0.256	1.794
		20	0.049	0.050	0.050	0.153	0.298	0.225	2.414
		10	0.034	0.012	0.023	0.041	0.113	0.077	5.710
60	5	15	0.019	0.002	0.011	0.011	0.098	0.055	8.144
		20	0.012	0	0.007	0.003	0.088	0.045	10.538
		10	0.012	0.001	0.007	0.004	0.031	0.017	12.360
90	5	15	0.007	0	0.004	0	0.023	0.011	17.276
		20	0.005	0	0.003	0	0.018	0.009	22.146
		10	0.002	0	0.001	0	0.007	0.004	20.097
120	5	15	0	0	0	0	0.004	0.002	27.579
		20	0	0	0	0	0.003	0.001	34.989

5 Example

We use the spherical data given in Wood (1982) (Table 2 in Schmidt, 1976) consisting of a set of 33 estimates of a previous magnetic pole position of the earth obtained using palaeomagnetic techniques. Each estimate is associated with a different site, the 33 sites being spread over a large area of Tasmania. Schmidt says that "the data appear to fall into two main groups which are derived from two distinct geographical regions" .

Table 5: Estimates of the misclassification probabilities and distance D between the two populations for $p = 3$

$\theta(^o)$	κ_1	κ_2	$\widehat{P(1\|2)}$	$\widehat{P(2\|1)}$	Average	D
		10	0.188	0.305	0.247	2.325
30	5	15	0.136	0.270	0.203	3.566
		20	0.107	0.240	0.174	5.260
		25	0.088	0.215	0.152	6.520
		10	0.084	0.117	0.101	6.750
60	5	15	0.053	0.094	0.074	9.658
		20	0.036	0.079	0.058	12.630
		25	0.026	0.068	0.047	15.590
		10	0.009	0.009	0.009	19.250
120	5	15	0.004	0.005	0.005	26.327
		20	0.002	0.003	0.003	33.380
		25	0.001	0.002	0.002	40.390
		10	0.001	0.002	0.002	25.500
180	5	15	0	0.001	0.001	34.660
		20	0	0	0	43.750
		25	0	0	0	52.800

In discriminant analysis we need to know *a priori* the groups of observations. The *EM* (Estimation-Maximisation) algorithm proposed by Dempster, Laird and Rubin (1977) can be used in many contexts and in particular, to estimate the parameters of a mixture of

Table 6: Estimates of the misclassification probabilities and distance D between the two populations for $p = 4$

$\theta(^o)$	κ_1	κ_2	$\widehat{P(1\|2)}$	$\widehat{P(2\|1)}$	Average	D
		10	0.204	0.287	0.246	1.066
30	5	15	0.147	0.238	0.193	3.878
		20	0.112	0.199	0.156	6.483
		25	0.088	0.171	0.130	7.486
		10	0.100	0.120	0.110	6.408
60	5	15	0.066	0.092	0.079	9.475
		20	0.046	0.072	0.059	15.171
		25	0.034	0.059	0.047	15.777
		10	0.014	0.013	0.014	17.868
120	5	15	0.007	0.007	0.007	24.770
		20	0.004	0.004	0.004	31.621
		25	0.003	0.003	0.003	38.470
		10	0.002	0.003	0.003	23.595
180	5	15	0.001	0.001	0.001	32.420
		20	0	0.001	0.001	41.125
		25	0	0	0	49.800

distributions, more precisely to solve the likelihood equations. This algorithm also enables us to obtain a partition of data into groups, which are required in discriminant analysis. So, we have used this algorithm in this example to obtain a partition of the data into two groups

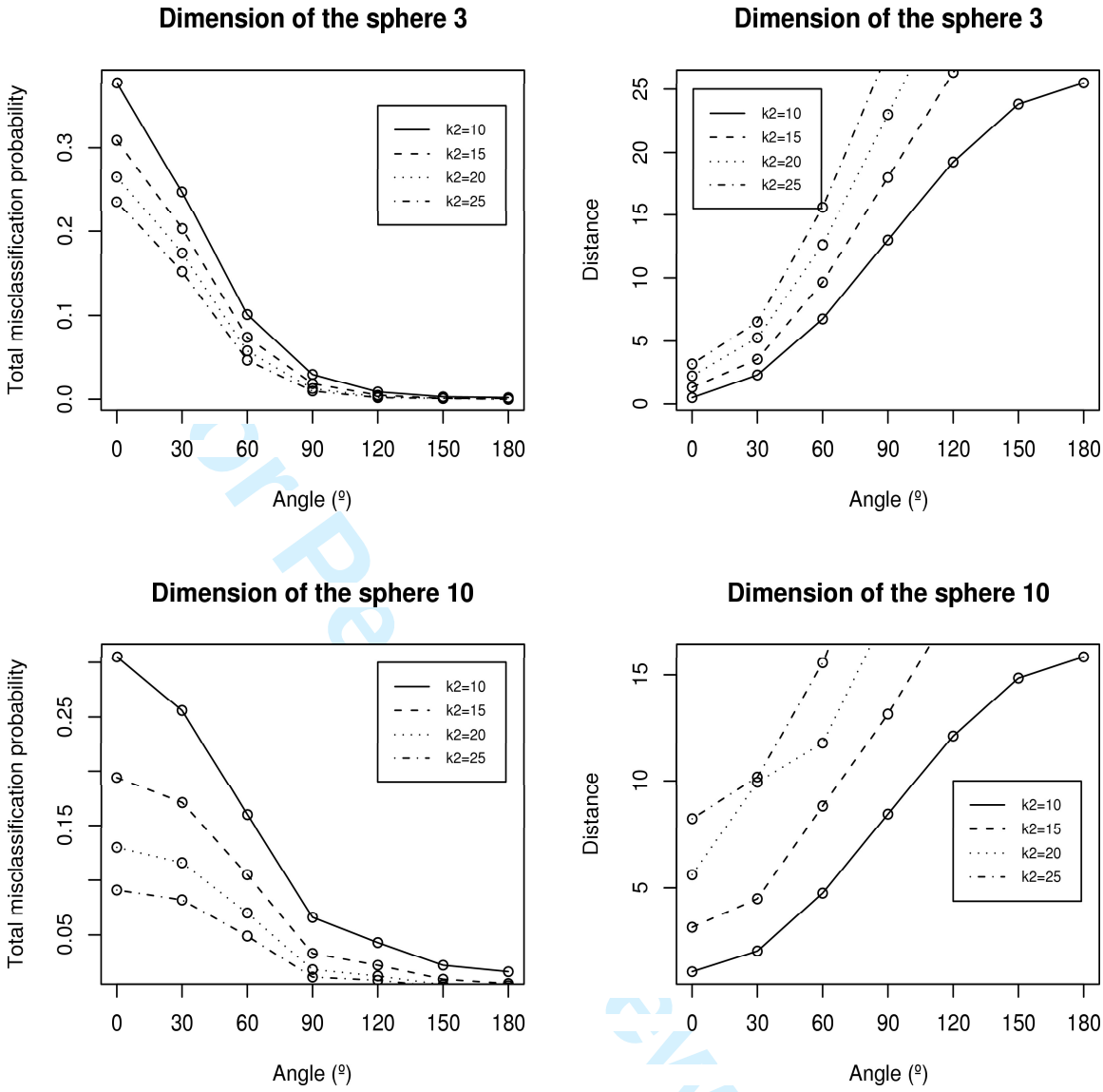

Figure 2: Estimate of the total misclassification probability and distance D for $\kappa_1 = 5$.

and also to estimate the parameters of the mixture of two Fisher distributions. This algorithm was already applied to these data in Figueiredo (2008). The partition of data obtained is the following: the group 1 contains the observations nos. 9,10,11,12,14,16,23,24,30 and the group 2 contain the remaining observations. We note that these two groups obtained

with the *EM* algorithm are the same groups suggested in Wood (1982).

Before applying discriminant analysis to these data, we check informally through chi--square Q-Q plots whether the two groups come from the Fisher distribution, because if an observation comes from a Fisher distribution with large concentration parameter, then the approximation (2.4) holds. We use in this approximation the maximum likelihood estimates because the parameters of the populations are unknown. These estimates are given by $\widehat{\boldsymbol{\mu}}_1 = (-0.318, 0.548, 0.736)'$, $\widehat{\kappa}_1 = 21.93$ and $\widehat{\boldsymbol{\mu}}_2 = (-0.623, 0.104, 0.775)'$, $\widehat{\kappa}_2 = 33.97$ and were also obtained by the *EM* algorithm. The chi-square Q-Q plots obtained for these data sets of size 9 and 24 are indicated in Fig.3. The fit to the Fisher distribution to each data set seems reasonable, and so for illustrating discriminant analysis, we have used these two samples.

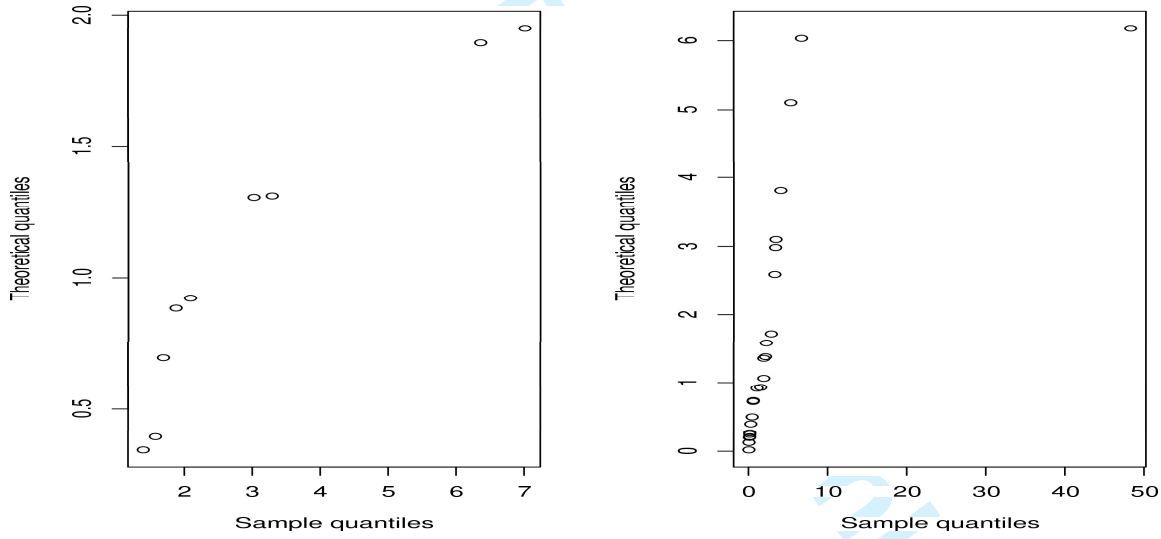

Figure 3: Chi-square Q-Q plots for the two samples

It is also important to check whether the two populations are distinct, before defining the discriminant rule. So, we have used the two-sample Watson-Williams test (Mardia and

Jupp, 2000, p. 219) which enables us to test the equality of the mean directions of the two populations. The F-statistic test is equal to 21.579 and so, comparing with the 95% percentage point of $F_{2,62}$ distribution, which is equal to 3.145, we reject the null hypothesis of equality of the two mean directions for the usual level of significance of 5%. As this test assumes that the concentration parameters of the two populations are equal, we test this hypothesis using the test given in Mardia and Jupp, 2000, p. 220-221. The F-statistic test is equal to 0.8119 and comparing with the 97.5% percentage point of $F_{16,46}$ distribution, which is equal to 2.106, we do not reject the null hypothesis of equality of the two concentration parameters for the usual level of significance of 5%.

In the discriminant rule defined by (3.4), we suppose equal costs of misclassification and the parameters of the populations estimated by their maximum likelihood estimates. We obtain the following discriminant function:

$$W_{12} = (\widehat{\kappa}_1 \widehat{\boldsymbol{\mu}}_1 - \widehat{\kappa}_2 \widehat{\boldsymbol{\mu}}_2)' \mathbf{x} + \ln \frac{c_p(\widehat{\kappa}_2)\, \widehat{\pi}_2}{c_p(\widehat{\kappa}_1)\, \widehat{\pi}_1},$$

where we consider the following estimates for the *prior* probabilities: $\widehat{\pi}_1 = 0.284$ and $\widehat{\pi}_2 = 0.716$ also obtained through the EM algorithm in Figueiredo (2008).

First, for validating the discriminant rule, we calculate the apparent error rate, obtained assigning the two samples using this discriminant rule. The apparent error rate is $1/33 \simeq 3\%$; only the observation no. 1 was assigned wrongly. This error rate is biased, as the samples used for obtaining the rule are assigned with the rule. So, next we determine the cross--validation error rate based on the "Leave-One-Out" (LOO) method proposed by Lachenbruch and Mickey (1968). The LOO method consists in the following: we remove one observation from one of the samples and based on the remaining observations we obtain the rule, which is used for assigning the removed observation; next, we remove another observation of the sample, and we determine the rule, which is used for assigning the removed

observation and we repeat the process until all observations of the two samples have been assigned. The error rate is the proportion of removed observations assigned incorrectly, and we obtained for this error rate $1/33 \simeq 3\%$. Again, the observation which is not assigned correctly was the observation no. 1. The distance D given by (3.7) with the parameters replaced by their maximum likelihood estimates is equal to 8.361. As the angle between the estimates of the mean directions is approximately 34^o, we can say that the value 8.361 is not large, because based on the simulation study (Sec. 4.2), we know that the distance between the two populations is not very large for a small angle (30^o) and this distance increases as the angle increases.

References

[1] Dempster, A. P.; Laird, N. M. and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society*, B, 3, 1-38.

[2] El Khattabi, S. and F. Streit (1996). Identification analysis in directional statistics. *Computational Statistics and Data Analysis* **23**, 45-63.

[3] Figueiredo, A. and P. Gomes (2006). Discriminant analysis based on the Watson distribution defined on the hypersphere. *Statistics* **40**, 5, 435-445.

[4] Figueiredo, A. (2008). Clustering estimates of a magnetic pole position of the earth. *Submitted for publication.*

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

[6] Fisher, N. I. (1993). *Statistical analysis of circular data*, Prentice Hall.

[7] Hill, G. W. (1977). Algorithm 518: Incomplete Bessel function I0(x); the von Mises distribution, *ACM Trans. Math. Software*, 3, 279-284.

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

[9] Johnson, R. A. and Wichern (1992). *Applied Multivariate Statistical Analysis*, 3rd ed, Cambridge University Press.

[10] Lachenbruch, P. A. and M. R. Mickey (1968). Estimation of error rates in Discriminant Analysis, *Technometrics*, 10, 1-11.

[11] Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*. John Wiley and Sons, Chichester.

[12] Morris, J. E. and P. J. Laycock (1974). Discriminant analysis of directional data. *Biometrika* **61**, 2, 335-341.

[13] Schmidt, P. W. (1976). The non-uniqueness of the Australian Mesozoic palaeomagnetic pole position. *Geophys. J. R. Astr. Soc.*, 47, 285-300.

[14] Stephens, M. A. (1982). Use of von Mises distribution to analyse continuous proportions. *Biometrika*, 69, **1**, 197-203.

[15] Stephens, M. A. (1992). On Watson's ANOVA for directions: In K.V. Mardia (Ed.). *The Art of Statistical Science*, 75-85 (John Wiley and Sons Ltd).

[16] Watson, G. S. (1983). *Statistics on spheres*. John Wiley and Sons, New York.

[17] Wood, A. (1982). A bimodal distribution on the sphere, *Applied Statistics*, **31**, 1, 52-58.

[18] Wood, A. (1994). Simulation of the von Mises-Fisher distribution. *Communications in Statistics - Simulation and Computation* **23**, 1, 157-164.