REVSTAT – Statistical Journal
Volume 4, Number 2, June 2006, 111–130

PERFORMANCE OF THE *EM* ALGORITHM ON THE IDENTIFICATION OF A MIXTURE OF WATSON DISTRIBUTIONS DEFINED ON THE HYPERSPHERE

Authors: Adelaide Figueiredo
– Faculdade de Economia da Universidade do Porto e LIACC,
Rua Dr. Roberto Frias, 4200-464 Porto, Portugal
Adelaide@fep.up.pt

Paulo Gomes
– Instituto Nacional de Estatística,
Av. António José Almeida, 1000-043 Lisboa, Portugal
Paulo.Gomes@ine.pt

Received: April 2004 Revised: August 2005 Accepted: March 2006

Abstract:

- We consider a set of n individuals described by p standardised variables, and we suppose that the individuals are previously selected from a population and the variables are a sample of variables assumed to come from a mixture of k bipolar Watson distributions defined on the hypersphere. In this context we provide the identification of the mixture through the *EM* algorithm and we also carry out a simulation study to compare the maximum likelihood estimates obtained from samples of moderate size with the respective asymptotic estimates. Our simulation results revealed good performance of the *EM* algorithm for moderate sample sizes.

Key-Words:

- *EM algorithm; mixture; principal components; Watson distribution.*

AMS Subject Classification:

- 62H11, 62H12, 62H25.

1. INTRODUCTION

We consider multivariate data with n individuals described by p variables. In the classical approach it is usual to assume that the p variables are fixed and the n individuals are randomly selected from a population of individuals. Now, we consider that the n individuals are fixed and the p variables are randomly selected from a population of variables. We standardise the variables to be points on the unit sphere in $\mathbb{R}^n$, denoted by $S_{n-1} = \{\mathbf{x} \in \mathbb{R}^n \colon \mathbf{x}'\mathbf{x} = 1\}$.

We suppose that the group of available variables on S_{n-1} is composed of k subgroups of variables and each subgroup comes from a bipolar Watson distribution. So we associate the sample of variables to a mixture of k bipolar Watson distributions defined on the hypersphere, as in Gomes [9]. This author considers an approach, based on the sampling of variables, and introduces some new results concerning the estimation of the parameters of the bipolar Watson distribution, taking into account not a sample of individuals but, a sample of variables. This type of ideas was referred to by Hotelling [10] who, in the context of Principal Components, studied the convergence of the eigenvalues and eigenvectors of the covariance matrix of groups of variables randomly chosen from a population of variables, when the dimension of the groups increases. Escoufier [5] also proposed a new coefficient for evaluating the proximity of two groups of variables, but supposing that the variables are observed.

For the identification of the mixture, we use the well-known *EM* algorithm proposed in Dempster, Laird and Rubin [3] (see Redner and Homer [14]).[1] This algorithm was developed to solve the likelihood equations in problems of incomplete data and we apply it to estimate the parameters of a mixture of k bipolar Watson distributions (see Figueiredo [7]).

The bipolar Watson distribution has been much used for axial data on the sphere (see Watson [16], Fisher, Lewis and Embleton [8] and Mardia and Jupp [13]). This distribution is denoted by $W_n(\mathbf{u},\xi)$ and it has density probability function given by

$$(1.1)\qquad f(\mathbf{x}) = \left\{{}_1F_1\left(\frac{1}{2},\frac{n}{2},\xi\right)\right\}^{-1} \exp\left\{\xi\left(\mathbf{u}'\mathbf{x}\right)^2\right\}, \quad \mathbf{x}\in S_{n-1},\;\; \mathbf{u}\in S_{n-1},\;\; \xi>0\,,$$

where the normalising constant is the reciprocal of a confluent hipergeometric function defined by

$$(1.2)\qquad {}_1F_1\left(\frac{1}{2},\frac{n}{2},\xi\right) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \int_0^1 e^{\xi t}\, t^{-0.5}(1-t)^{(n-3)/2}\, dt\;.$$

[1]Another possible method for the identification of the mixture is the k-means method proposed in Diday and Schroeder [15] (see Gomes [9]).

This distribution has two parameters: a directional parameter $\mathbf{u}$ and a concentration parameter ξ, which measures the concentration about $\pm\mathbf{u}$. As ξ increases, the distribution becomes more concentrated about $\pm\mathbf{u}$. This is a rotationally symmetric distribution about the principal axis $\pm\mathbf{u}$ and it is bimodal, with modes $\mathbf{u}$ and $-\mathbf{u}$.

Let $X = \left[\mathbf{x}^1|\mathbf{x}^2|...|\mathbf{x}^p\right]$ be a random sample of variables from the bipolar Watson distribution $W_n(\mathbf{u},\xi)$. The maximum likelihood estimator of $\mathbf{u}$ is the eigenvector associated with the largest eigenvalue $\widehat{w}$ of $XX' = \sum_{i=1}^{p} \mathbf{x}^i\,\mathbf{x}^{i\,\prime}$, that is, $\widehat{\mathbf{u}}$ is defined by $(XX')\,\widehat{\mathbf{u}} = \widehat{w}\,\widehat{\mathbf{u}}$. So, it follows that the maximum likelihood estimator of the directional parameter $\mathbf{u}$ based on the sample of variables is the first principal component of the sample. The maximum likelihood estimator of ξ is the solution of the equation $Y(\widehat{\xi}) = \widehat{w}/p$, where the function $Y(\xi)$ is defined by $Y(\xi) = \frac{d}{d\xi} \ln {}_1F_1(1/2, n/2, \xi)$.

The estimators $\widehat{\xi}$ and $\widehat{w}$ have asymptotic Gaussian distribution (see Gomes [9] and Bingham [1]):

$$\widehat{\xi} \;\dot\sim\; N\left(\xi, \frac{1}{p\,Y_{11}^2(\xi)}\right) \qquad \text{and} \qquad \frac{\widehat{w}}{p} \;\dot\sim\; N\left(Y(\xi), \frac{Y_{11}^2(\xi)}{p}\right). \tag{1.3}$$

where the function $Y_{11}^2(\xi)$ is defined by $Y_{11}^2(\xi) = \frac{d^2}{d\xi^2} \ln {}_1F_1(\frac{1}{2}, \frac{n}{2}, \xi)$.

In this study we consider the particular case of a bipolar Watson distribution. If we had assumed $\xi < 0$ in (1.1), we would obtain a girdle Watson distribution and the study of this distribution would be similar to the one that is done in this paper.

In Section 2 we present the identification of the mixture of k bipolar Watson distributions through the *EM* algorithm. In Section 3 we carry out a simulation study to compare the behaviour of the estimators obtained through the *EM* algorithm for moderate samples with the respective asymptotic estimators. In Section 4 we give some concluding remarks.

2. IDENTIFICATION OF A MIXTURE OF *k* BIPOLAR WATSON DISTRIBUTIONS DEFINED ON THE HYPERSPHERE

The density function of a mixture of k bipolar Watson components $C_1, ..., C_k$ defined on the hypersphere, whose identifiability was proved by Kent [12], is given by

$$\begin{aligned} g(\mathbf{x}|\phi) = \sum_{j=1}^{k} \pi_j f(\mathbf{x}|\theta_j)\,, \qquad & \mathbf{x}\in S_{n-1},\;\; 0<\pi_j<1,\;\; j=1,...,k,\;\; \sum_{j=1}^{k}\pi_j = 1\,, \\ & \phi = \big(\mathbf{u}_1,...,\mathbf{u}_k,\xi_1,...,\xi_k,\pi_1,...,\pi_k\big), \quad \theta_j = (\mathbf{u}_j,\xi_j)\;, \end{aligned} \tag{2.1}$$

where $(\pi_1, ..., \pi_k)$ are the proportions of the mixture and $f(\mathbf{x}|\theta_j)$ is the density function corresponding to the C_j component.

As a mixture of distributions may be seen as a problem of incomplete data (see Everitt and Hand [6]), the *EM* algorithm may be applied to solve the likelihood equations in the estimation of the parameters of a mixture of k bipolar Watson distributions.

Let $\left[\mathbf{x}^1|\mathbf{x}^2|...|\mathbf{x}^p\right]$ be a random sample from the mixture and let $Z = \left[\mathbf{z}_1|...|\mathbf{z}_p\right]$ be the missing data, where the indicator vector $\mathbf{z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ik})$ with $Z_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}^i \in C_j \\ 0 & \text{if } \mathbf{x}^i \notin C_j \end{cases}$, $\sum\limits_{j=1}^{k} Z_{ij} = 1$ indicates the component of the variable $\mathbf{x}^i$ of the mixture.

The log likelihood associated with the complete sample $\left[\mathbf{x}^1|...|\mathbf{x}^p|Z\right]$ is given by

$$L\big(\phi|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\, \ln\Big\{\pi_j f\big(\mathbf{x}^i|\theta_j\big)\Big\}\,, \tag{2.2}$$

where $t_j(\mathbf{x}^i)$ is the *posterior* probability of $\mathbf{x}^i$ belonging to C_j defined by

$$t_j(\mathbf{x}^i) \;=\; \frac{\pi_j\, f(\mathbf{x}^i|\theta_j)}{\sum\limits_{h=1}^{k} \pi_h\, f(\mathbf{x}^i|\theta_h)}\;. \tag{2.3}$$

The log likelihood associated with the complete sample given by (2.2) may be written as

$$L\big(\phi|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) \;=\; L\big(\phi_1|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) + L\big(\phi_2|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big)\;,$$

where

$$L\big(\phi_1|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\, \ln f(\mathbf{x}^i|\theta_j)\,, \qquad \phi_1 = (\theta_1, ..., \theta_k)$$

and

$$L\big(\phi_2|\mathbf{x}^1, ..., \mathbf{x}^p, Z\big) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j(\mathbf{x}^i)\, \ln \pi_j\,, \qquad \phi_2 = (\pi_1, ..., \pi_k)\;.$$

To estimate the vector of unknown parameters ϕ of the mixture, the *EM* algorithm proceeds iteratively in two steps:

$$E - \text{Estimation} \quad \text{and} \quad M - \text{Maximisation}\,.$$

The algorithm starts with the initial solution:

$$\phi^0 \;=\; \big(\mathbf{u}_1^0, ..., \mathbf{u}_k^0, \xi_1^0, ..., \xi_k^0, \pi_1^0, ..., \pi_k^0\big)\;.$$

In the m-th iteration, the two steps are:

***E*-Step**

Use estimates $\phi^{(m)}$ of the parameters of the mixture in the m-th iteration for $j=1,...,k$ and $i=1,...,p$ to estimate the *posterior* probability of $\mathbf{x}^i$ belonging to the j-th component of the mixture

$$t_j^{(m)}(\mathbf{x}^i) \;=\; \frac{\pi_j^{(m)} f\big(\mathbf{x}^i|\theta_j^{(m)}\big)}{\sum\limits_{h=1}^{k} \pi_h^{(m)} f\big(\mathbf{x}^i|\theta_h^{(m)}\big)} \;. \tag{2.4}$$

***M*-Step**

Use estimates $t_j^{(m)}(\mathbf{x}^i)$ to maximise the logarithm of the likelihood function $L(\phi_1|\mathbf{x}^1,...,\mathbf{x}^p,Z)$.

First, we consider the function $L(\phi_1)$, subject to the constraint $\mathbf{u}_j'\mathbf{u}_j = 1$:

$$L(\phi_1) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\Big[-\ln\big\{{}_1F_1(1/2,n/2,\xi_j)\big\} + \xi_j(\mathbf{u}_j'\mathbf{x}^i)^2\Big] - \lambda_1(\mathbf{u}_j'\mathbf{u}_j - 1) \;,$$

where λ_1 is a Lagrange multiplier and $t_j^{(m)}(\mathbf{x}^i)$ is defined in (2.4).

The maximum likelihood estimate of $\mathbf{u}_j$ is the solution of the following equation:

$$\frac{\partial L(\phi_1)}{\partial \mathbf{u}_j} \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, 2\,\xi_j\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\mathbf{u}_j - 2\,\lambda_1\mathbf{u}_j \;=\; 0 \;. \tag{2.5}$$

We premultiply the last expression by $\mathbf{u}_j\prime$ to obtain

$$\lambda_1 \;=\; \xi_j \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{u}_j\prime\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\mathbf{u}_j \;.$$

Then, the maximum likelihood estimator of $\mathbf{u}_j\prime$ in the $(m{+}1)$-th iteration, $\widehat{\mathbf{u}}_j^{(m+1)}$ is the eigenvector associated with the eigenvalue $\widehat{w}_j$, that is

$$\left(\sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\right)\widehat{\mathbf{u}}_j^{(m+1)} \;=\; \widehat{w}_j\, \widehat{\mathbf{u}}_j^{(m+1)} \;, \qquad j = 1,..,k \;, \tag{2.6}$$

where $\widehat{w}_j$ is a eigenvalue of $\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \mathbf{x}^i\mathbf{x}^{i\,\prime}$ and it is given by

$$\widehat{w}_j \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\, \widehat{\mathbf{u}}_j^{(m+1)}\prime\, \mathbf{x}^i\mathbf{x}^{i\,\prime}\, \widehat{\mathbf{u}}_j^{(m+1)} \;.$$

Next, we show that we maximise $L(\phi_1)$ if we consider the largest eigenvalue of the matrix. In fact, the function $L(\phi_1)$ can be written in the form

$$L(\phi_1) \;=\; -\sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln\bigl\{{}_1F_1(1/2,n/2,\xi_j)\bigr\} + \sum_{j=1}^{k}\xi_j\,\widehat{w}_j \;.$$

As $\ln{}_1F_1(1/2,n/2,\xi_j) > 0$, we have $\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln{}_1F_1(1/2,n/2,\xi_j) > 0$. We also have $\widehat{w}_j \geq 0$ because $\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,\mathbf{x}^i\mathbf{x}^{i\,\prime}$ is a positive definite matrix. Consequently, the function $L(\phi_1)$ is maximised if $\widehat{w}_j$ is maximum.

Second, the maximum likelihood estimator of ξ_j is the solution of the following equation

$$\frac{\partial L(\phi_1)}{\partial \xi_j} \;=\; \sum_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,\Bigl\{-Y(\xi_j) + (\mathbf{u}_j\prime\mathbf{x}^i)^2\Bigr\} \;=\; 0 \;,$$

where the function $Y(.)$ is defined in Section 1. The solution of this equation leads to the maximum of $L(\phi_1)$ as we show that $\partial^2 L(\phi_1)/\partial\xi_j^2 < 0$, $\forall_{\xi_j}$. In fact, $\partial^2 L(\phi_1)/\partial\xi_j^2 = -\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)\,dY(\xi_j)/\xi_j$ and $Y(\xi)$ is an increasing function (see Gomes [9]).

Then, the maximum likelihood estimator of ξ_j in the $(m{+}1)$-th iteration, $\widehat{\xi}_j^{(m+1)}$, is the solution of the equation

$$Y\bigl(\widehat{\xi}_j^{(m+1)}\bigr) \;=\; \frac{\widehat{w}_j}{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)} \;, \qquad j = 1,...,k \;. \tag{2.7}$$

Third, we consider the function $L(\phi_2)$, subject to the constraint $\sum\limits_{j=1}^{k}\pi_j = 1$:

$$L(\phi_2) \;=\; \sum_{i=1}^{p}\sum_{j=1}^{k} t_j^{(m)}(\mathbf{x}^i)\;\ln\pi_j \;-\; \lambda_2\biggl(\sum_{j=1}^{k}\pi_j - 1\biggr) \,,$$

where λ_2 is a Lagrange multiplier. The maximum likelihood estimator of π_j is the solution of the following equation

$$\frac{\partial L(\phi_2)}{\partial \pi_j} \;=\; \frac{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)}{\pi_j} - \lambda_2 \;=\; 0 \;.$$

We sum the last equation for j from 1 to k to obtain $\lambda_2 = p$. Then, the maximum likelihood estimator of π_j in the $(m{+}1)$-th iteration, $\widehat{\pi}_j^{(m+1)}$ is given by

$$\widehat{\pi}_j^{(m+1)} \;=\; \frac{\sum\limits_{i=1}^{p} t_j^{(m)}(\mathbf{x}^i)}{p} \;, \qquad j = 1,...,k \;. \tag{2.8}$$

The estimation of the parameters $\mathbf{u}_j$ and ξ_j associated with the j-th component gives us a privileged direction as well as a measure of dispersion of the j-th cluster around this direction.

A partition $(P_1, ..., P_k)$ of the sample of variables is obtained assigning the variable $\mathbf{x}^j$ to the component for which the *posterior* probability is the largest, that is,

$$P_j = \left\{ \mathbf{x}^i : \; t_j(\mathbf{x}^i) = \max_h t_h(\mathbf{x}^i), \;\; h=1, ..., k \right\} \tag{2.9}$$

and when $t_j(\mathbf{x}^i) = t_h(\mathbf{x}^i)$ consider $\mathbf{x}^i \in P_j$ if $j < h$.

3. SIMULATION STUDY

We considered a mixture with equal proportions ($\pi_1 = \pi_2 = 0.5$) of two bipolar Watson distributions: $W_n(\mathbf{u}_1, \xi_1)$ and $W_n(\mathbf{u}_2, \xi_2)$, with $\xi_1 = \xi_2 = \xi$, $\mathbf{u}_1 = (0, ..., 0, 1)$ and $\mathbf{u}_2 = (0, ...0, (1-\cos^2\theta)^{1/2}, \cos\theta)$, where θ is the angle between $\mathbf{u}_1$ and $\mathbf{u}_2$. The bipolar Watson distribution is rotationally symmetric about the directional parameter, so if we had used, for each θ, other directional parameters $\mathbf{u}_1$ and $\mathbf{u}_2$, we would have obtained the same results in our study. For the simulation of the bipolar Watson distribution we used a rejection-type method (see Huo [11] and Bingham [2]). We considered two dimensions of the sphere $n = 10{,}30$. For each n, we assumed equal samples size $p_1 = p_2 = p = 30(10)100$, several values of the concentration parameter $\xi = 10(10)50{,}100$ and several values of the angle $\theta = 18°, 54°, 90°$. For each case, we considered *2500* replicates of the *EM* algorithm. In each replicate, we used a randomly chosen initial solution and a sufficiently large number of iterations (*100*) to obtain the final solution. We supposed that the algorithm converged, in a certain replicate, if the condition:

$$\left| \Big(L\big(\phi^{(m+1)}\big) - L\big(\phi^{(m)}\big) \Big) \Big/ L\big(\phi^{(m+1)}\big) \right| \leq 10^{-5}$$

holds in the last five iterations, where $L(\phi^{(m)})$ denotes the likelihood of the sample in the m-th iteration. For each n and p, the *EM* algorithm converged in most part of the replicates, it did not converge only in very few replicates when ξ is very small or θ is small.

In each replicate we determined the following estimates $\widehat{\xi}_j$, $\widehat{w}_j/p_j$, $j=1,2$, $\widehat{\theta}$, $\widehat{\pi}_j$, $j=1,2$ of the parameters ξ_j, $Y(\xi_j)$, $j=1,2$, θ, π_j, $j=1,2$, respectively, where p_j is the dimension of the j-th group, which is equal to $\sum_{i=1}^{p} t_j(\mathbf{x}^i)$. Then, we calculated the average and the standard deviation of the estimates obtained in all replicates, denoted by $\overline{\widehat{\xi}}_j$, $\overline{\widehat{w}}_j/p_j$, $j=1,2$, $\overline{\widehat{\theta}}$, $\overline{\widehat{\pi}}_j$, $j=1,2$ and $s(\widehat{\xi}_j)$, $s(\widehat{w}_j/p_j)$, $j=1,2$, $s(\widehat{\theta})$, $s(\widehat{\pi}_j)$, $j=1,2$, respectively. If in a replicate the *EM* algorithm did not converge we excluded that replicate for calculating the average and the standard deviation of the estimates.

By (1.3) the asymptotic expected value of $\widehat{\xi}_j$ and $\widehat{w}_j/p_j$ are ξ_j and $Y(\xi_j)$ respectively, $j=1,2$. In Table 1 and Figure 1, we indicate the values of $Y(\xi)$ [2] for each n and ξ.

Table 1: Values of $Y(\xi)$ for each n and ξ.

$n \diagdown \xi$	10	20	30	40	50	60	70	80	90	100
10	0.500	0.766	0.847	0.886	0.909	0.924	0.935	0.943	0.950	0.955
30	0.074	0.241	0.496	0.630	0.706	0.756	0.791	0.817	0.838	0.854

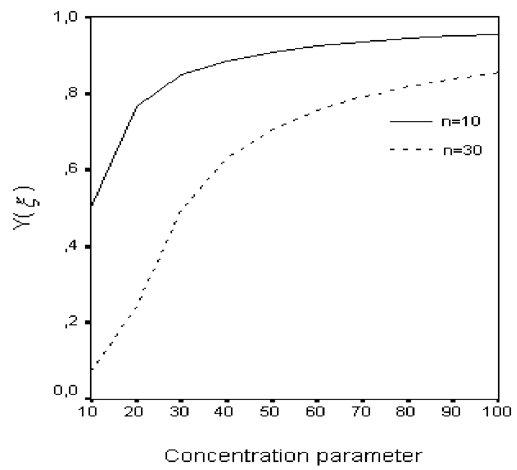

Figure 1: Values of $Y(\xi)$ for $n=10$ and $n=30$.

As expected for each n, $Y(\xi)$ is an increasing function with ξ, which tends to *1*, when ξ increases (see Gomes [9], p. 43–45). For each ξ, the function $Y(\xi)$ increases when n decreases.

We determined the estimated relative bias of the estimators given by the expressions: $(\overline{\widehat{\xi}}_j-\xi_j)/\xi_j$, $(\overline{\widehat{w}}_j/p_j-Y(\xi_j))/Y(\xi_j)$, $j=1,2$, $(\overline{\widehat{\theta}}-\theta)/\theta$, $(\overline{\widehat{\pi}}_j-\pi_j)/\pi_j$, $j=1{,}2$ and the estimated mean squared error (MSE) given by: $s^2(\widehat{\xi}_j)+(\overline{\widehat{\xi}}_j-\xi_j)^2$, $s^2(\widehat{w}_j/p_j)+(\overline{\widehat{w}}_j/p_j-Y(\xi_j))^2$, $j=1,2$, $s^2(\widehat{\theta})+(\overline{\widehat{\theta}}-\theta)^2$, $s^2(\widehat{\pi}_j)+(\overline{\widehat{\pi}}_j-\pi_j)^2$, $j=1,2$.

[2] We obtained the function $Y(\xi)$ using the Kummer function, which is defined by $M(a,b,z) = 1+\sum_{i=1}^{\infty}\big\{a(a+1)...(a+i-1)z^i\big\}/\big\{b(b+1)...(b+i-1)\,i!\big\}$ or by the integral $M(a,b,z) = \Gamma(b)/\big\{\Gamma(b-a)\Gamma(a)\big\}\int_0^1 e^{zt}\,t^{a-1}(1-t)^{b-a-1}dt$, where ${}_1F_1(1/2,n/2,\xi)=M(1/2,n/2,\xi)$.

We indicate the results of our simulation study in the Tables A1–A4 of the Appendix and in the Figures 2–8. In the tables of the Appendix, the algorithm converged in all replicates for each case. We have also produced another 4 tables, which were not included: two tables for the relative bias (for $n = 10$ and $n = 30$) and two tables for the *MSE* (for $n = 10$ and $n = 30$) of the estimators when the concentration parameter ξ varies.

In Figure 2 we observe that

- As expected, the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ are asymptotically unbiased, that is the estimated relative bias of these estimators tends to *0* as the sample size p increases. For fixed ξ and p, the relative bias of $\widehat{\xi}_1$ and $\widehat{\xi}_2$ tends to decrease when θ increases. For an angle $\theta = 90°$ or $\theta = 54°$, the bias of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ is relatively small and when $\theta = 90°$ the bias is not greater than 10% of the true value of the concentration parameter (for $n = 10{,}30$, $\xi = 30{,}100$ and $p = 30(10)100$).

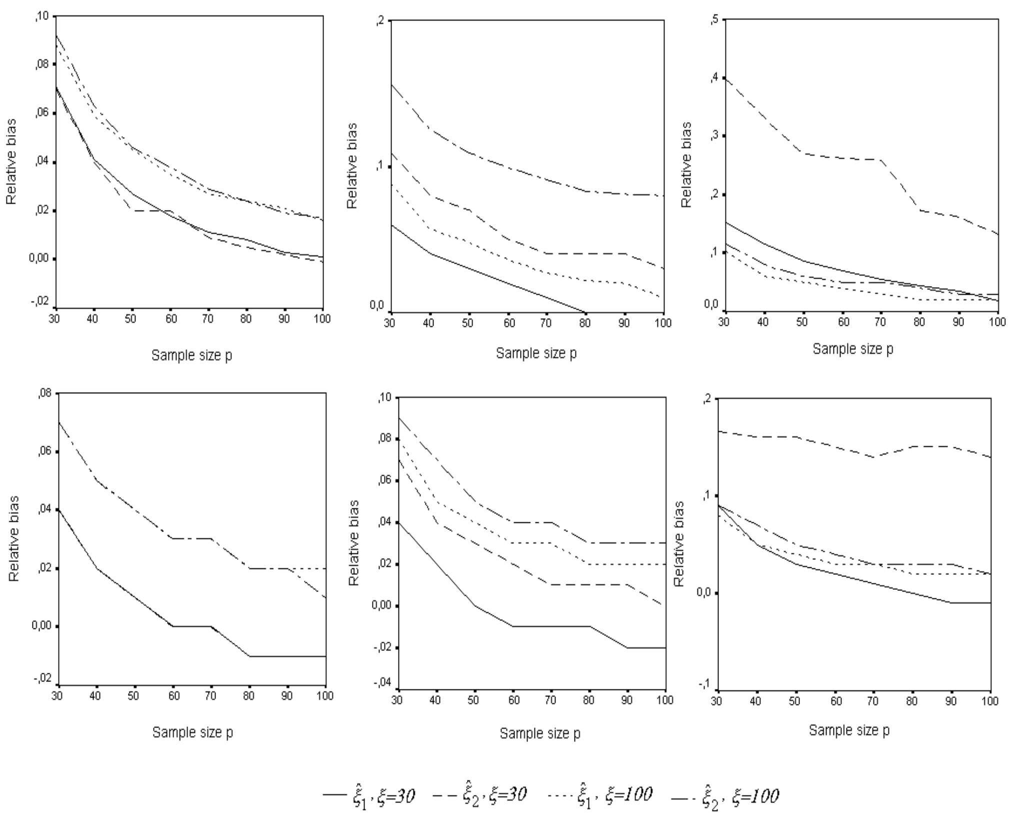

Figure 2: Relative bias of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when p varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 3 we observe that

- As expected, in general the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ become more efficient as p increases. When the angle is large or moderate ($\theta = 90°$ or $\theta = 54°$) and $\xi = 30$, these estimators have relatively small *MSE* and become less efficient when ξ increases.

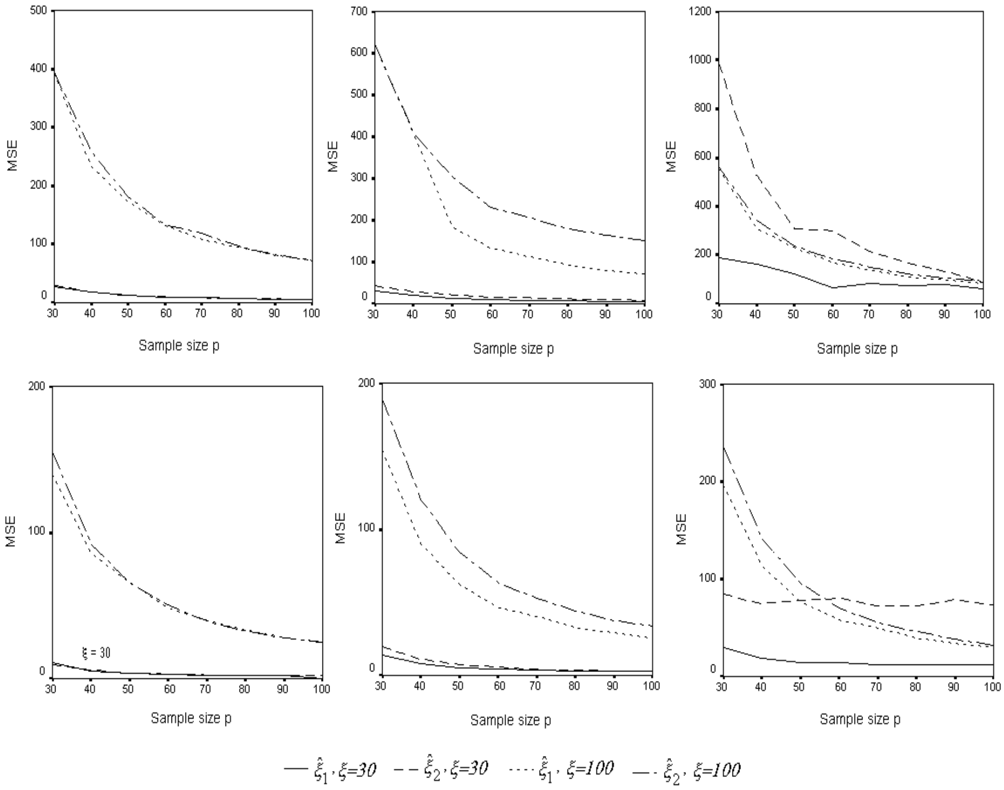

Figure 3: Mean squared error of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when p varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 4 we observe that

- When the angle is moderate or large ($\theta = 54°$ or $\theta = 90°$), the bias of $\widehat{\xi}_1$ and $\widehat{\xi}_2$ is very small and maintains approximately constant or increases slightly as ξ increases for $\xi \geq 20$ when $n = 10$ and for $\xi \geq 30$ when $n = 30$. When $n = 10$ and $\theta = 18°$, the bias of the estimators is relatively large, but it decreases when ξ increases.

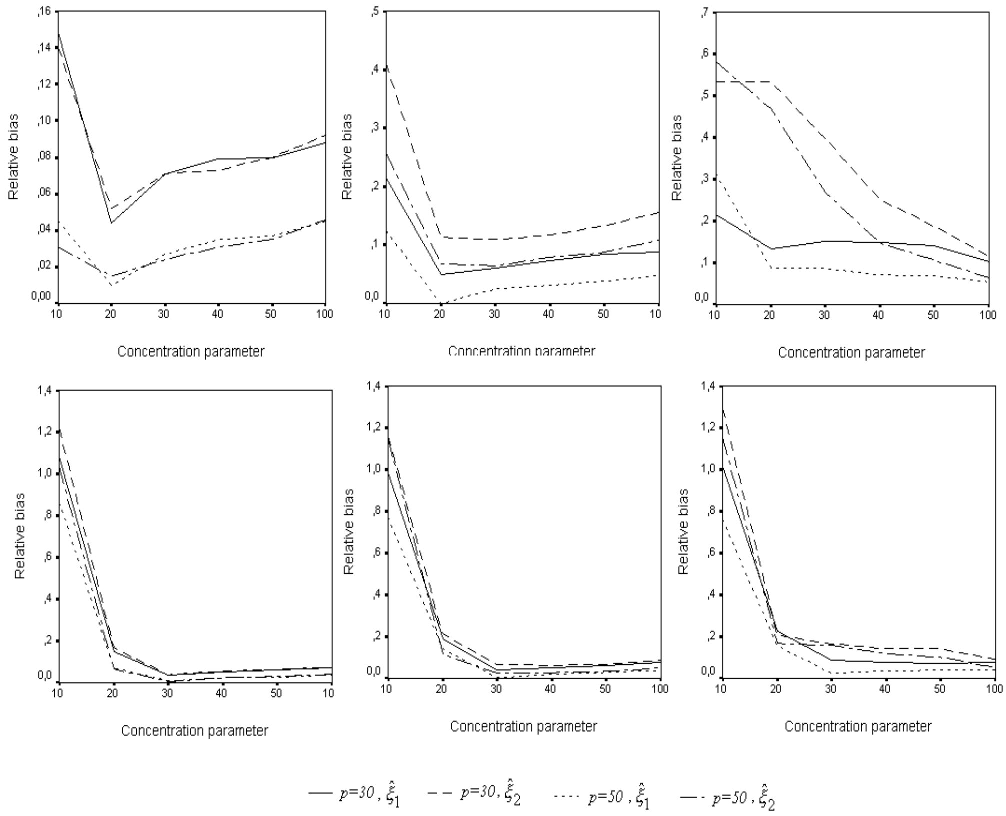

Figure 4: Relative bias of the estimators $\widehat{\xi_1}$ and $\widehat{\xi_2}$ when ξ varies (in top: $n = 10$, in bottom: $n = 30$ and from left to right: angle $90°$, $54°$, $18°$).

In Figure 5 we observe that

- When the angle is moderate or large ($\theta = 54°$ or $\theta = 90°$), the *MSE* of the estimators $\widehat{\xi_1}$ and $\widehat{\xi_2}$ increases when ξ increases for $\xi \geq 30$ and so these estimators become less efficient.
- The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are unbiased or have very small bias for every p and ξ. When $\theta = 90°$ the bias of these estimators is not greater than approximately 3% of the respective parameter. The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are asymptotically unbiased, that is, the estimated relative bias of the estimators tends to *0* as the sample size p increases. See Tables A1–A2 of the Appendix.

In Figure 6 we observe that

- The estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ have bias approximately equal to *0* for $\xi \geq 20$ when $n = 10$ and for $\xi \geq 30$ when $n = 30$.
- As the *MSE* of the estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ are *0* or approximately *0*, these estimators are very efficient. See Tables A3–A4 of the Appendix.

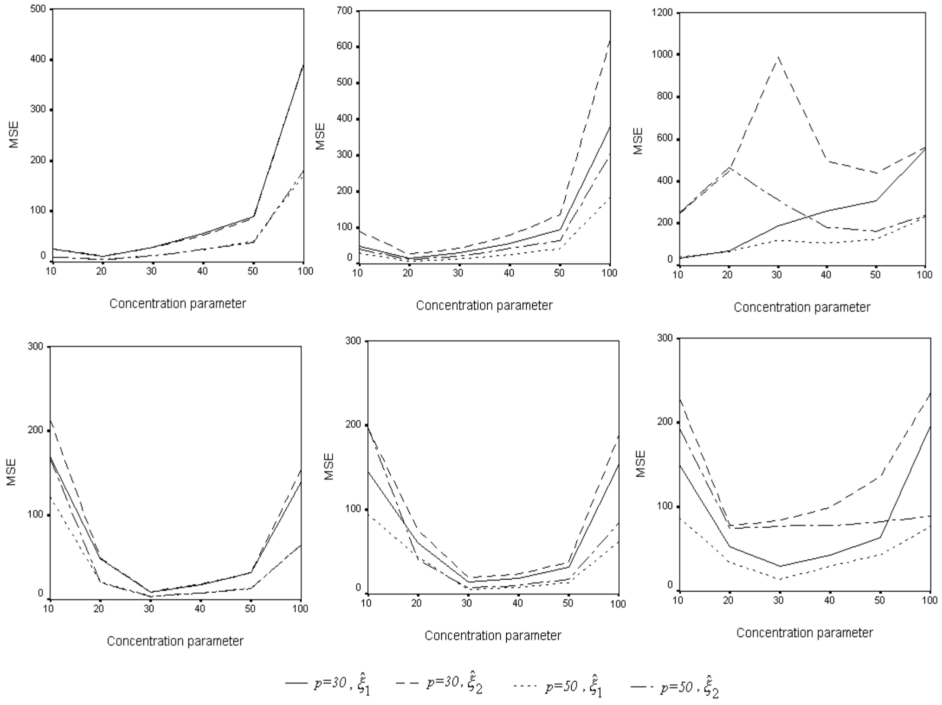

Figure 5: Mean squared error of the estimators $\widehat{\xi}_1$ and $\widehat{\xi}_2$ when ξ varies (in top: $n=10$, in bottom: $n=30$ and from left to right: angle 90°, 54°, 18°).

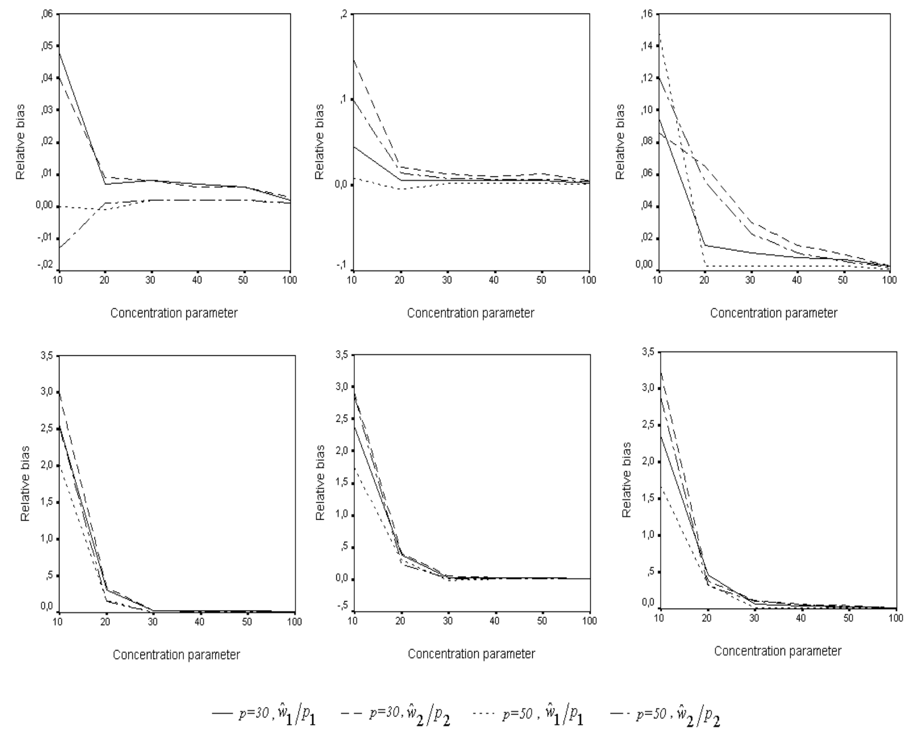

Figure 6: Relative bias of the estimators $\widehat{w}_1/p_1$ and $\widehat{w}_2/p_2$ when ξ varies (in top: $n=10$, in bottom: $n=30$ and from left to right: angle 90°, 54°, 18°).

In Figure 7 we observe that

- The estimator $\widehat{\theta}$ has relatively small *MSE*, except for $n = 30$ and $\xi = 30$ when the relative bias and the standard deviation of $\widehat{\theta}$ are relatively large. The *MSE* of the estimator $\widehat{\theta}$ decreases when p increases.

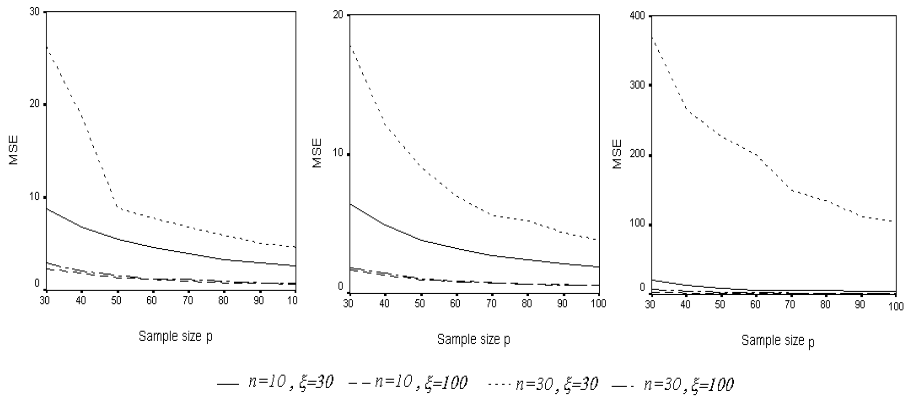

Figure 7: Mean squared error of the estimator $\widehat{\theta}$ when p varies (from left to right: angle 90°, 54°, 18°).

In Figure 8 we observe that

- For every θ and $\xi \geq 20$, the *MSE* of the estimator $\widehat{\theta}$ decreases when ξ increases.
- The estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are unbiased or present very small bias for the analysed cases, except in some cases when $\theta = 18°$. See Tables A1–A2 of the Appendix.
- The estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ have *MSE* equal to *0* or approximately *0*, and so these estimators are very efficient. See Tables A3–A4 of the Appendix.

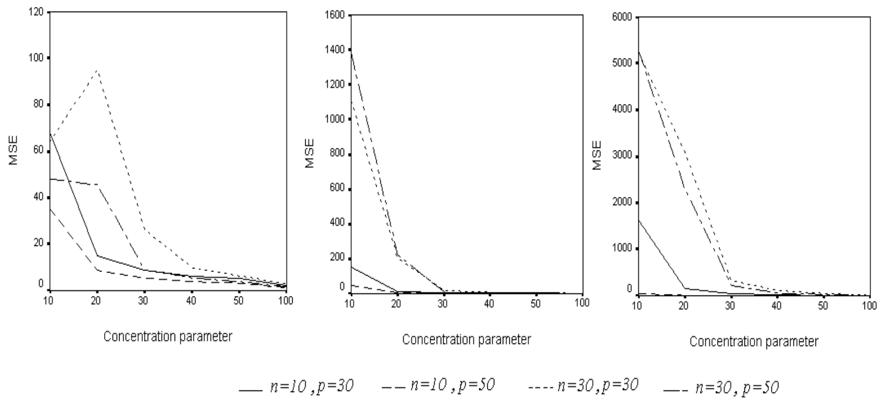

Figure 8: Mean squared error of the estimator $\widehat{\theta}$ when ξ varies (from left to right: angle 90°, 54°, 18°).

4. CONCLUSION

The simulation study has revealed a good identification of a mixture of bipolar Watson distributions defined on the hypersphere through the *EM* algorithm.

The performance of this algorithm is good for moderate sample sizes, essentially on the estimation of the prior probabilities and on the estimation of the directional parameters of the mixture. For a large or moderate angle θ between the directional parameters of the mixture, the efficiency of the estimators of the concentration parameters of the mixture is better for moderate values (neither very small nor very large) of the true concentration parameters. The estimation of the angle θ is very efficient in general and the efficiency of $\widehat{\theta}$ improves as the concentration parameter increases.

REFERENCES

[1] Bingham, C. (1964). *Distribution on the sphere and on the projective plane*, Ph. D. Thesis, Yale University.

[2] Bingham, C. (1983). *Simulation on the n-sphere: The Bingham distribution*, Rapport Technique, University of Minnesota.

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

[4] Diday, E. (1972). *Nouveaux concepts et nouvelles méthodes en classification automatique*, Thèse de Doctorat d'État, Université Paris VI.

[5] Escoufier, Y. (1968). Le traitement des variables vectorielles, *Biometrics*, **29**, 31–47.

[6] Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*, Chapman and Hall, Ltd, London.

[7] Figueiredo, A. (2000). *Classificação de variáveis no contexto de um modelo probabilístico definido na n-esfera*, Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa.

[8] Fisher, N.I.; Lewis, T. and Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data*, Cambridge University Press.

[9] Gomes, P. (1987). *Distribution de Bingham sur la n-sphere: une nouvelle approche de l'Analyse Factorielle*, Thèse d'État, Université des Sciences et Techniques du Languedoc-Montpellier.

[10] Hotteling, H. (1933). Analysis of a complex of statistical variables into principal components, *J. Educational Psychology*, **24**, 417–441.

[11] Kent, J.T. (1983). Identifiability of finite mixtures for directional data, *Annals of Statistics*, **11**, 984–988.

[12] Huo, V. (1984). *Small samples from Bingham distributions*, Ph. D. thesis, University of Minnesota.

[13] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*, 2nd Ed., John Wiley and Sons, Chichester.

[14] Redner, R. and Homer, W. (1984). Mixture Densities, Maximum Likelihood, EM algorithm, *Journal of the American Statistical Association*, **B**, 39, 1–38.

[15] Schroeder, A. (1981). Analyse d'un mélange de distributions de probabilité de même type, *Revue de Statistique Appliquée*, **XXIV**, 1, 39–62.

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

APPENDIX

Table 2: Relative bias of the estimators for $n=10$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	0.071	0.071	0.008	0.008	0.000	0.005	−0.005
		40	0.041	0.041	0.004	0.004	−0.001	−0.001	0.001
		50	0.027	0.024	0.002	0.002	−0.001	0.004	−0.004
		60	0.018	0.017	0.001	0.001	0.000	−0.005	0.005
		70	0.011	0.009	0.000	0.000	0.000	0.001	−0.001
		80	0.008	0.005	0.000	−0.001	0.001	0.001	−0.001
		90	0.003	0.002	−0.001	−0.001	0.000	0.000	0.000
		100	0.001	−0.001	−0.001	−0.002	0.002	0.001	−0.001
	54°	30	0.061	0.109	0.006	0.013	−0.002	0.002	−0.002
		40	0.042	0.080	0.004	0.010	−0.002	0.001	−0.001
		50	0.025	0.065	0.002	0.008	−0.004	0.000	0.000
		60	0.020	0.051	0.001	0.006	−0.005	−0.002	0.002
		70	0.011	0.044	0.000	0.005	−0.006	0.000	0.000
		80	0.002	0.040	−0.001	0.005	−0.007	0.000	0.000
		90	0.002	0.037	−0.001	0.005	−0.007	−0.001	0.001
		100	−0.004	0.033	−0.002	0.005	−0.007	−0.001	0.001
	18°	30	0.152	0.398	0.011	0.030	0.256	0.116	−0.116
		40	0.116	0.331	0.006	0.027	0.192	0.086	−0.086
		50	0.086	0.269	0.003	0.023	0.157	0.070	−0.070
		60	0.070	0.262	0.004	0.022	0.125	0.061	−0.061
		70	0.055	0.259	0.000	0.018	0.107	0.049	−0.050
		80	0.044	0.173	−0.001	0.016	0.086	0.030	−0.030
		90	0.034	0.161	−0.002	0.016	0.077	0.031	−0.031
		100	0.018	0.132	−0.004	0.014	0.068	0.020	−0.020
100	90°	30	0.088	0.092	0.002	0.003	0.000	0.000	0.000
		40	0.059	0.063	0.002	0.002	0.000	−0.003	0.003
		50	0.045	0.046	0.001	0.001	0.000	0.003	−0.003
		60	0.035	0.038	0.001	0.001	0.000	0.001	−0.001
		70	0.027	0.029	0.001	0.001	0.000	0.000	0.000
		80	0.024	0.024	0.000	0.000	0.000	0.000	0.000
		90	0.021	0.019	0.000	0.000	0.000	0.002	−0.002
		100	0.016	0.017	0.000	0.000	0.000	−0.001	0.001
	54°	30	0.088	0.156	0.002	0.005	−0.003	0.000	0.000
		40	0.057	0.125	0.001	0.004	−0.003	0.002	−0.002
		50	0.048	0.109	0.001	0.004	−0.004	−0.001	0.001
		60	0.036	0.099	0.001	0.003	−0.004	0.000	0.000
		70	0.027	0.091	0.001	0.003	−0.004	0.001	−0.001
		80	0.022	0.083	0.000	0.003	−0.005	0.000	0.000
		90	0.018	0.081	0.000	0.003	−0.005	−0.001	0.001
		100	0.014	0.080	0.000	0.003	−0.005	−0.001	0.001
	18°	30	0.103	0.116	0.003	0.003	0.041	−0.001	0.001
		40	0.064	0.081	0.002	0.002	0.032	0.000	0.000
		50	0.054	0.064	0.001	0.002	0.027	−0.002	0.002
		60	0.040	0.052	0.001	0.001	0.023	−0.001	0.001
		70	0.029	0.046	0.001	0.001	0.021	0.001	−0.001
		80	0.024	0.037	0.000	0.001	0.019	0.000	0.000
		90	0.020	0.034	0.000	0.001	0.018	−0.002	0.002
		100	0.015	0.032	0.000	0.001	0.017	−0.002	0.002

Table 3: Relative bias of the estimators for $n = 30$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	0.036	0.038	0.029	0.031	−0.001	0.001	−0.001
		40	0.017	0.018	0.012	0.012	0.000	0.000	0.000
		50	0.009	0.007	0.005	0.003	0.001	0.003	−0.003
		60	0.001	0.000	−0.004	−0.005	0.000	0.004	−0.004
		70	−0.004	−0.002	−0.008	−0.006	0.000	−0.001	0.001
		80	−0.006	−0.008	−0.011	−0.020	0.001	−0.005	0.005
		90	−0.011	−0.010	−0.015	−0.014	0.000	0.001	−0.001
		100	−0.011	−0.013	−0.016	−0.017	0.000	0.001	−0.001
	54°	30	0.040	0.067	0.028	0.053	0.048	0.009	−0.009
		40	0.017	0.043	0.008	0.035	0.038	0.008	−0.008
		50	0.003	0.026	−0.012	0.019	0.031	0.006	−0.006
		60	−0.005	0.019	−0.012	0.013	0.025	0.007	−0.007
		70	−0.010	0.011	−0.017	0.007	0.022	0.006	−0.006
		80	−0.014	0.007	−0.021	0.002	0.021	0.006	−0.006
		90	−0.018	0.005	−0.025	0.001	0.017	0.006	−0.006
		100	−0.020	0.001	−0.027	−0.003	0.016	0.004	−0.004
	18°	30	0.085	0.166	0.065	0.113	0.994	0.218	−0.218
		40	0.049	0.156	0.033	0.106	0.857	0.243	−0.243
		50	0.025	0.157	0.011	0.106	0.773	0.274	−0.274
		60	0.020	0.153	0.006	0.100	0.700	0.274	−0.274
		70	0.009	0.144	−0.006	0.096	0.629	0.286	−0.286
		80	−0.003	0.148	−0.018	0.100	0.581	0.292	−0.292
		90	−0.005	0.151	−0.021	0.100	0.531	0.290	−0.290
		100	−0.006	0.140	−0.023	0.092	0.493	0.275	−0.275
100	90°	30	0.070	0.074	0.010	0.011	0.000	0.001	−0.001
		40	0.049	0.053	0.007	0.008	−0.001	0.005	−0.005
		50	0.038	0.039	0.006	0.006	0.000	0.003	−0.003
		60	0.030	0.031	0.004	0.005	0.000	−0.001	0.001
		70	0.026	0.025	0.004	0.004	0.000	0.001	−0.001
		80	0.020	0.021	0.003	0.003	0.000	0.003	−0.003
		90	0.018	0.017	0.003	0.003	0.000	−0.002	0.002
		100	0.015	0.014	0.002	0.002	0.000	0.000	0.000
	54°	30	0.075	0.087	0.011	0.012	0.003	0.004	−0.004
		40	0.051	0.065	0.008	0.009	0.002	−0.004	0.004
		50	0.039	0.052	0.006	0.008	0.001	0.002	−0.002
		60	0.029	0.042	0.004	0.006	−0.001	0.002	−0.002
		70	0.025	0.036	0.004	0.006	−0.001	0.002	−0.002
		80	0.021	0.033	0.003	0.005	−0.002	0.000	0.000
		90	0.018	0.028	0.003	0.004	−0.002	0.000	0.000
		100	0.016	0.027	0.002	0.004	−0.002	−0.002	0.002
	18°	30	0.078	0.091	0.011	0.013	0.119	0.006	−0.007
		40	0.054	0.067	0.008	0.010	0.095	−0.005	0.005
		50	0.040	0.052	0.006	0.008	0.080	0.001	−0.001
		60	0.030	0.041	0.004	0.006	0.066	0.001	−0.001
		70	0.026	0.034	0.004	0.005	0.059	0.001	−0.001
		80	0.021	0.030	0.003	0.005	0.052	−0.002	0.002
		90	0.016	0.025	0.002	0.004	0.047	0.001	−0.001
		100	0.015	0.022	0.002	0.003	0.044	−0.004	0.004

Table 4: Mean squared error of the estimators for $n = 10$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	28.39	29.34	0.001	0.001	8.8	0.01	0.01
		40	17.49	18.51	0	0	6.81	0.01	0.01
		50	12.99	12.57	0	0	5.47	0.01	0.01
		60	10.00	10.14	0	0	4.67	0	0
		70	8.41	8.52	0	0	3.92	0	0
		80	7.14	6.90	0	0	3.25	0	0
		90	6.40	6.53	0	0	2.92	0	0
		100	5.77	5.47	0	0	2.67	0	0
	54°	30	29.61	42.68	0.001	0.001	6.43	0.01	0.01
		40	20.20	29.07	0	0.001	4.89	0.01	0.01
		50	13.51	21.59	0	0.001	3.80	0	0
		60	11.28	16.18	0	0.001	3.22	0	0
		70	8.94	14.29	0	0.001	2.70	0	0
		80	7.32	11.97	0	0.001	2.40	0	0
		90	6.38	9.97	0	0.001	2.09	0	0
		100	5.87	8.59	0	0.001	1.94	0	0
	18°	30	188.76	987.30	0.001	0.002	34.48	0.04	0.04
		40	163.08	526.42	0.001	0.002	22.30	0.03	0.03
		50	120.85	309.95	0.001	0.002	16.37	0.03	0.03
		60	66.77	299.22	0.001	0.002	10.97	0.03	0.03
		70	83.52	214.79	0.001	0.001	9.52	0.03	0.03
		80	75.39	167.57	0.001	0.001	7.45	0.02	0.02
		90	77.56	133.97	0.001	0.001	6.14	0.02	0.02
		100	47.93	86.04	0.001	0.001	5.42	0.02	0.02
100	90°	30	391.95	393.74	0	0	2.34	0.01	0.01
		40	233.19	258.01	0	0	1.82	0.01	0.01
		50	172.30	181.35	0	0	1.37	0	0
		60	132.61	133.23	0	0	1.17	0	0
		70	108.59	118.57	0	0	0.98	0	0
		80	94.51	96.45	0	0	0.81	0	0
		90	81.51	82.20	0	0	0.81	0	0
		100	71.97	72.44	0	0	0.66	0	0
	54°	30	380.48	620.53	0	0	1.70	0.01	0.01
		40	239.14	407.87	0	0	1.30	0.01	0.01
		50	183.84	304.88	0	0	0.98	0.01	0.01
		60	140.48	263.35	0	0	0.88	0	0
		70	112.76	209.09	0	0	0.79	0	0
		80	93.25	179.25	0	0	0.64	0	0
		90	79.89	164.52	0	0	0.64	0	0
		100	71.10	150.27	0	0	0.56	0	0
	18°	30	554.58	560.82	0	0	2.79	0.01	0.01
		40	309.45	342.75	0	0	2.07	0.01	0.01
		50	231.62	238.22	0	0	1.51	0.01	0.01
		60	169.17	186.41	0	0	1.27	0	0
		70	138.87	151.15	0	0	1.13	0	0
		80	110.62	121.23	0	0	0.92	0	0
		90	97.90	106.03	0	0	0.82	0	0
		100	83.85	92.72	0	0	0.77	0	0

Table 5: Mean squared error of the estimators for $n = 30$ with the sample size p

ξ	θ	p	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$\widehat{w}_1/p_1$	$\widehat{w}_2/p_2$	$\widehat{\theta}$	$\widehat{\pi}_1$	$\widehat{\pi}_2$
30	90°	30	8.99	9.33	0.002	0.002	26.27	0.01	0.01
		40	5.36	5.82	0.002	0.002	18.66	0.01	0.01
		50	3.83	3.78	0.001	0.001	8.85	0	0
		60	3.03	3.13	0.001	0.001	7.78	0	0
		70	2.52	2.52	0.001	0.001	6.81	0	0
		80	2.17	2.10	0.001	0.001	5.92	0	0
		90	1.97	1.96	0.001	0.001	5.06	0	0
		100	1.74	1.85	0.001	0.001	4.67	0	0
	54°	30	13.87	19.54	0.004	0.004	17.86	0.01	0.01
		40	8.14	11.07	0.002	0.003	12.06	0.01	0.01
		50	5.24	7.29	0.002	0.002	9.08	0.01	0.01
		60	4.14	5.72	0.001	0.002	6.97	0	0
		70	3.48	4.12	0.001	0.001	5.58	0	0
		80	3.02	3.55	0.001	0.001	5.23	0	0
		90	2.92	3.19	0.001	0.001	4.33	0	0
		100	2.61	2.73	0.001	0.001	3.83	0	0
	18°	30	29.40	84.02	0.005	0.011	370.63	0.03	0.03
		40	19.16	74.94	0.004	0.011	265.15	0.03	0.03
		50	14.18	77.88	0.003	0.011	226.23	0.04	0.04
		60	14.27	80.57	0.003	0.011	200.45	0.04	0.04
		70	12.62	72.60	0.003	0.011	148.84	0.04	0.04
		80	11.84	72.02	0.003	0.010	134.46	0.05	0.05
		90	11.80	78.76	0.003	0.011	112.28	0.23	0.23
		100	12.62	73.23	0.003	0.010	104.27	0.05	0.05
100	90°	30	139.69	154.52	0	0	2.92	0.01	0.01
		40	85.91	92.13	0	0	2.08	0.01	0.01
		50	65.74	65.13	0	0	1.59	0.01	0.01
		60	48.48	50.47	0	0	1.17	0	0
		70	39.77	39.20	0	0	1.17	0	0
		80	33.22	32.60	0	0	0.98	0	0
		90	28.12	27.72	0	0	0.81	0	0
		100	25.01	24.67	0	0	0.81	0	0
	54°	30	154.06	189.17	0	0	1.86	0	0
		40	89.81	120.39	0	0	1.43	0.01	0.01
		50	62.23	84.41	0	0	1.06	0.01	0.01
		60	46.53	63.51	0	0	0.84	0	0
		70	40.38	52.90	0	0	0.75	0	0
		80	32.60	44.35	0	0	0.67	0	0
		90	29.15	37.53	0	0	0.58	0	0
		100	25.36	33.66	0	0	0.58	0	0
	18°	30	196.39	234.80	0	0	7.18	0.01	0.01
		40	113.25	141.25	0	0	4.67	0.01	0.01
		50	76.84	96.11	0	0	3.39	0.01	0.01
		60	58.06	69.951	0	0	2.57	0	0
		70	49.88	55.35	0	0	2.15	0	0
		80	39.48	45.93	0	0	1.74	0	0
		90	33.79	38.03	0	0	1.50	0	0
		100	30.11	32.01	0	0	1.37	0	0