Journal of Mathematical Extension
Vol. XX, No. XX, (2014), pp-pp (Will be inserted by layout editor)
ISSN: 1735-8299
URL: http://www.ijmex.com

Nonsingular Information Matrix and the von Mises Fisher Distribution

Maryam Ghodsi
Department of Mathematics and Statistics, Jahrom Branch, Islamic Azad University, Jahrom, Iran

Abstract. In this article, the asymptotic distribution of the deviance statistic for some hypothesis tests concerning the parameters of the von Mises Fisher distribution is discussed. The focus is on the likelihood of the distribution. We find the distribution of the deviance statistic using Chernoff's idea about the distance from the cone constructed from the null and alternative hypotheses.

AMS Subject Classification: 62E20; 62H11; 62H15
Keywords and Phrases: Asymptotic distribution theory, Directional data, Hypothesis testing

1 Introduction

To analyse big data, finding the asymptotic distribution of the deviance statistic is of great help. Chernoff [2] introduced a method for finding the asymptotic distribution of this statistic in some nonstandard situations. His methodology then was extended by others such as Self & Liang [9], Feder [3], Moran [8], Chant [1], Geyer [5] and Vu & Zhou [11]. Silvapulle & Sen [10] expand the methodology in more detail and continue to write a book considering constrained statistical analysis.

Received: XXXX; Accepted: XXXX (Will be inserted by editor)

In this paper, we derive the asymptotic distribution of deviance statistics for some hypothesis tests concerning the parameters of the von Mises Fisher distribution.

Watson [12], Fisher, Embleton & Lewis [4] and Mardia & Jupp [7] treat the problem by taking a geometrical view point of this distribution. Our work is mainly towards the likelihood function of the distribution, considering the first and second derivative of the log likelihood function. In a standard setup, the likelihood should satisfy some conditions which are not true with the von Mises Fisher distribution [6]. However, if we eliminate one parameter and replace it by a function of the other parameters, then it satisfies some standard conditions as we show in Section 3. Referring to Vu & Zhou [11] for a full multivariate setup, we check the assumptions and in Section 4, the asymptotic distribution of the deviance statistics for some suggested hypotheses are presented.

As is shown in [6], eliminating one parameter is not however a good method to analyse the von Mises Fisher distribution, because it results in an unintuitive and complicated expressions. It is preferable to develop a direct methodology for this distribution which can be found in [6].

The von Mises Fisher distribution is used in [6] to analyse high dimensional asset allocation of financial portfolios built from various stock indices of countries. The methodology that I use in this paper is different from the one in my thesis.

von Mises Fisher distribution is one type of spherical distributions whose relatively informative and simple formulation makes it useful in the study of the directional data. Directional data locates data on a circle, sphere or hypersphere. Typical examples of directional data are related to the earth and celestial sphere. The data can be represented by a vector $\mathbf{x}$ which satisfy the condition $\mathbf{x}^T\mathbf{x} = 1$.

Let $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^d : x_1^2 + x_2^2 + \cdots + x_d^2 = 1\}$ denote a unit sphere in $\mathbb{R}^d$. The von Mises Fisher distribution is defined on this sphere by the following density function:

$$f_{\mathbf{X}}(\mathbf{x}) = c(\kappa)^{-1} \exp\{\kappa \boldsymbol{\mu}^T \mathbf{x}\}, \quad \mathbf{x} \in \mathbb{S}^d, \kappa \geq 0, \boldsymbol{\mu} \in \mathbb{S}^d, \tag{1}$$

where $f_{\mathbf{X}}(x_1, x_2, \cdots, x_d)$ is the density of a d dimensional random vector $\mathbf{X}$ at the point $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ on the surface of the sphere. In this distribution $\kappa \geq 0$ represents the concentration and $\boldsymbol{\mu}$ is the mean direction or the pole such that $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_d)$ with $\boldsymbol{\mu} \in S^d$. κ is a measure of precision. If $\kappa = 0$, then the data are distributed uniformly over the sphere. When κ is large, the distribution is concentrated on a small portion of the sphere. $\boldsymbol{\mu}$ is called the modal or mean vector of the distribution and locates the density on the sphere.

In this investigation, we assume $\kappa \neq 0$, therefore we choose the parameter space to be $\Theta = \{(\kappa, \mu_1, \cdots, \mu_{d-1}) \in (0, \infty) \times (-1, 1)^{d-1}\}$ and the true value of the parameter to be $\boldsymbol{\theta}_0 = (\kappa_0, \mu_{10}, \cdots, \mu_{d-10}) \in \Theta$.

The first derivative of $c(\kappa)$ (the normalising constant in the von Mises Fisher distribution) with respect to κ will be denoted by c_κ and the second derivative by $c_{\kappa\kappa}$. These notations will be used throughout this paper.

2 Chernoff's innovation for the asymptotic distribution of the deviance statistic

Chernoff [2] extends the work of Wilks [13] by considering subsets of Θ such as hyperplanes, i.e., subspaces of dimension $d-1$ or less. The hyperplanes in $\mathbb{R}$ are points, in $\mathbb{R}^2$ are lines and in $\mathbb{R}^3$ are planes. Every hyperplane divides the space in two parts. Under some regularity conditions Chernoff considers the hypotheses

$$\begin{cases} H_0 & : \boldsymbol{\theta} \text{ is on one side of a hyperplane} \\ H_A & : otherwise, \end{cases}$$

so his emphasis is quite different from Wilks'. He locates the true value of the parameter, $\boldsymbol{\theta}_0$, on the boundary of the two disjoint subsets defined by the null and the alternative hypotheses. In the one-dimensional case the null and alternative hypotheses are simply

$$\begin{cases} H_0 & : \theta \geq \theta_0 \\ H_A & : \theta < \theta_0. \end{cases}$$

Let $\mathcal{N}$ be a neighborhood of $\boldsymbol{\theta}_0$ in Θ and let $\mathbf{X}$ be a random variable in $\mathbb{R}^d$ with density $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathcal{N}$. Chernoff assumes
(CH1) for almost all $\mathbf{x}$, the first, second and third derivatives of

$\log(f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}$ exist, for every $\boldsymbol{\theta} \in \mathcal{N}$;
(CH2) if $\boldsymbol{\theta} \in \mathcal{N}$, all of the first, second and third derivatives of $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ are bounded by finitely integrable functions where these functions are the same for the first and second derivatives and the expectation of the third one does not depend on $\boldsymbol{\theta}$.
(CH3) if $\boldsymbol{\theta} \in \mathcal{N}$, the matrix $\mathbf{S}_{\boldsymbol{\theta}}$ in (3) is finite and positive definite.

Throughout his proofs, he translates the origin so that $\boldsymbol{\theta}_0$ is zero and considers the hypotheses

$$\begin{cases} H_0 & : \boldsymbol{\theta} \in \Omega \subset \mathcal{N} \\ H_A & : \boldsymbol{\theta} \in \tau \subset \mathcal{N} \end{cases}$$

and illustrates his method of testing them through three examples which can be found in Chernoff [2], Silvapulle & Sen [10] and more explicitly in [6].

Recall that a set $C \subseteq \mathbb{R}^d$ is a cone with vertex at 0 if $\boldsymbol{\theta} \in C$ implies $a\boldsymbol{\theta} \in C$ for all $a > 0$. Chernoff introduces the idea of a set $\phi \subset \mathcal{N}$ approximated by the cone C_ϕ at $\mathbf{0}$ if

$$\inf_{\mathbf{x} \in C_\phi} ||\mathbf{x} - \mathbf{y}|| = o(||\mathbf{y}||) \quad \text{for } \mathbf{y} \in \phi$$

and

$$\inf_{\mathbf{y} \in \phi} ||\mathbf{x} - \mathbf{y}|| = o(||\mathbf{x}||) \quad \text{for } \mathbf{x} \in C_\phi,$$

then proves the following theorem:
Suppose $\boldsymbol{\theta}_0 = \mathbf{0}$ and $\widehat{\boldsymbol{\theta}}_\phi$ is the maximum likelihood estimator in a set $\phi \subset \mathcal{N}$ and
1) the regularity conditions (CH1), (CH2), (CH3) are satisfied,
2) the origin is a boundary point of ϕ implies that $\widehat{\boldsymbol{\theta}}_\phi \xrightarrow{\text{P}} 0$, for any $\phi \subset \mathcal{N}$,
3) the sets Ω and τ are approximated by nonnull and disjoint cones C_Ω and C_τ.
Then the asymptotic distribution of the deviance statistic

$$d_n = 2[\mathcal{L}_n(\hat{\boldsymbol{\theta}}_n^2) - \mathcal{L}_n(\hat{\boldsymbol{\theta}}_n^1)], \tag{2}$$

where $\hat{\boldsymbol{\theta}}_n^2$ and $\hat{\boldsymbol{\theta}}_n^1$ are local maxima of the log likelihood function $\mathcal{L}_n(\boldsymbol{\theta})$ on Ω and τ, is the same as it would be for the test of $\boldsymbol{\theta} \in C_\Omega$ against $\boldsymbol{\theta} \in C_\tau$ based on one observation from a normal distribution with mean $\mathbf{0}$ and variance $\mathbf{J}^{-1}$. In this setup $\mathbf{J} = \mathbf{S_0}$ with

$$\mathbf{S}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]; \quad \boldsymbol{\theta} \in \Theta, \tag{3}$$

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathrm{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})\right]; \quad \boldsymbol{\theta} \in \Theta. \tag{4}$$

3 Assumptions on the first and second derivative of the log likelihood function

Let $\Theta = \{(\kappa, \mu_1, \ldots, \mu_{d-1}), \kappa > 0, \mu_i \in (0,1)\}$ denote the parameter space, and $\boldsymbol{\theta}_0 = (\kappa_0, \mu_{10}, \ldots, \mu_{d-10})$ the true value of $\boldsymbol{\theta}$.

To analyze the asymptotic behavior of the MLEs, we need some assumptions on the asymptotic behavior of the first and second derivative matrices and their expectations. Before defining them, we calculate first, second and the expectations of the log likelihood function. We define the derivative of $\mathcal{L}_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ to be the d-vector

$$\mathbf{S}_n(\boldsymbol{\theta}) = [\frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \theta_2}, \ldots, \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \theta_d}]^T.$$

Similarly, we define the negative of the second derivative of $\mathcal{L}_n(\boldsymbol{\theta})$ to be the $d \times d$ symmetric matrix

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

Define $\mathbf{D}_n = E\{\mathbf{S}_n(\boldsymbol{\theta}_0)\mathbf{S}_n^T(\boldsymbol{\theta}_0)\}$ and $\mathbf{G}_n = E\{\mathbf{F}_n(\boldsymbol{\theta}_0)\}$, and

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_d),$$

so that

$$\mu_1 \overline{X}_1 + \mu_2 \overline{X}_2 + \ldots + \mu_d \overline{X}_d = \boldsymbol{\mu}^T \overline{\mathbf{X}}_n,$$

and

$$\mu_d\overline{X}_i - \mu_i\overline{X}_d = \widetilde{\boldsymbol{\mu}}^T\overline{\mathbf{X}}_{i,d}.$$

The first derivative of $\mathcal{L}_n$ is the $d \times 1$ vector

$$\mathbf{S}_n(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \kappa} \\ \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \mu_1} \\ \vdots \\ \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \mu_{d-1}} \end{bmatrix} = n \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \frac{\kappa}{\mu_d} I \end{bmatrix} \mathbf{M} \left(\overline{\mathbf{X}} - E\overline{\mathbf{X}} \right), \tag{5}$$

where matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{bmatrix} \mu_1 & \mu_d & 0 & \cdots & 0 \\ \mu_2 & 0 & \mu_d & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{d-1} & 0 & 0 & \cdots & \mu_d \\ \mu_d & -\mu_1 & -\mu_2 & \cdots & -\mu_{d-1} \end{bmatrix}_{d\times d} = \begin{bmatrix} \mathbf{a} & \mu_d \mathbf{I}_{(d-1)\times(d-1)} \\ \mu_d & -\mathbf{a}^T \end{bmatrix}, \tag{6}$$

and

$$\mathbf{a}^T = (\mu_1, \mu_2, \cdots, \mu_{d-1}), \tag{7}$$

It is easily verified that $E(\mathbf{S}_n(\boldsymbol{\theta}_0)) = 0$. The matrix $\mathbf{D}_n = E(\mathbf{S}_n(\boldsymbol{\theta}_0)\mathbf{S}_n^T(\boldsymbol{\theta}_0))$ is

$$\mathbf{D}_n = n^2 \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \frac{\kappa}{\mu_d} I \end{bmatrix} \mathbf{M}\{\mathrm{Var}(\overline{\mathbf{X}})\}\mathbf{M}^T \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \frac{\kappa}{\mu_d} I \end{bmatrix}.$$

Using the expectation formulae presented and proved in [6] (because it is assumed that (X_{i1}, X_{i2}) are n random samples) we obtain

$$\mathbf{D}_n = n\frac{\kappa c_\kappa}{\mu_d^2 c(\kappa)} \begin{bmatrix} \left(\frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}\right) / \frac{\kappa c_\kappa}{\mu_d^2 c(\kappa)} & \mathbf{0}^T \\ \mathbf{0} & F \end{bmatrix}, \tag{8}$$

where F is

$$F = \left(\mu_d^2 \mathbf{I} + \mathbf{a}\mathbf{a}^T\right). \tag{9}$$

After some computation, $\mathbf{F}_n(\boldsymbol{\theta})$ can be written as

$$\mathbf{F}_n(\boldsymbol{\theta}) = n \begin{bmatrix} \frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)} & \frac{-1}{\mu_d}\mathbf{b}^T \\ \frac{-1}{\mu_d}\mathbf{b} & \frac{\kappa}{\mu_d^3}\overline{X}_d F \end{bmatrix}, \tag{10}$$

where F is in (9), and

$$\mathbf{b}^T = \begin{bmatrix} \tilde{\boldsymbol{\mu}}_1^T \overline{\mathbf{X}}_{1,d} & \tilde{\boldsymbol{\mu}}_2^T \overline{\mathbf{X}}_{2,d} & \dots & \tilde{\boldsymbol{\mu}}_{d-1}^T \overline{\mathbf{X}}_{d-1,d} \end{bmatrix}. \tag{11}$$

Again using the moment formulae in [6], we see that the expectation of this matrix in $\boldsymbol{\theta}_0$, which is $\mathbf{G}_n$, is equal to $\mathbf{D}_n$.

Theorem 3.1. $\mathbf{D}_n = \mathbf{G}_n$ *is a nonsingular positive definite square matrix.*

Proof. Let $\mathbf{t}_d^T = [t_1, t_2, \dots, t_d]$ and $\mathbf{t}_{d-1}^T = [t_2, \dots, t_d]$ be two nonzero vectors respectively in $\mathbb{R}^d$ and $\mathbb{R}^{d-1}$,

$$\begin{aligned}
\mathbf{t}_d^T \mathbf{D}_n \mathbf{t}_d &= \frac{n}{\mu_d^2}\left(\frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}\right) t_1^2 + \left(n\frac{\kappa c_\kappa}{\mu_d^2 c(\kappa)}\right) \mathbf{t}_{d-1}^T F \mathbf{t}_{d-1} \\
&= \frac{n}{\mu_d^2}\left(\frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}\right) t_1^2 \\
&+ \left(n\frac{\kappa c_\kappa}{c(\kappa)}\right)(\mathbf{t}_{d-1}^T I \mathbf{t}_{d-1}) + \left(n\frac{\kappa c_\kappa}{\mu_d^2 c(\kappa)}\right) \mathbf{t}_{d-1}^T \mathbf{a}\mathbf{a}^T \mathbf{t}_{d-1} \\
&= \frac{n}{\mu_d^2}\left(\frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}\right) t_1^2 + \left(n\frac{\kappa c_\kappa}{c(\kappa)}\right)\sum_{i=2}^{d-1} t_i^2 + \left(n\frac{\kappa c_\kappa}{\mu_d^2 c(\kappa)}\right)\left(\sum_{i=2}^{d-1} t_i \mu_i\right)^2 \\
&> 0
\end{aligned}$$

From Theorem 2.1 of my thesis, we have $c_\kappa \geq 0$. Furthermore, if $c_\kappa = 0$, then $\mathbf{D}_n$ is zero, so as a result we assume that $c_\kappa > 0$. By remembering that $|\mu_d| \in (0, 1)$, we have the last expression positive just by applying the Cauchy-Schwarz inequality as below

$$\begin{aligned}
\left(\int_{|\mathbf{x}|=1} (\boldsymbol{\mu}^T\mathbf{x}) \exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\} dS\right)^2 \leq& \left(\int_{|\mathbf{x}|=1} (\boldsymbol{\mu}^T\mathbf{x})^2 \exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\} dS\right) \\
&\times \left(\int_{|\mathbf{x}|=1} \exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\} dS\right),
\end{aligned}$$

that is,

$$c_\kappa^2 \leq c_{\kappa\kappa} c(\kappa).$$

Further, we have strict inequality here because both the functions $\boldsymbol{\mu}^T \mathbf{x}$ and 1 are nonzero and non-proportional through a scalar. So we have

$$c_\kappa^2 < c_{\kappa\kappa} c(\kappa) \to \frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)} > 0$$

and the proof ends. $\square$

For any positive definite matrix C, let $C^{1/2}(C^{T/2})$ be a left (the corresponding right), square root of C, that is, any matrices satisfying $C^{1/2}C^{T/2} = C$, where $C^{T/2} = (C^{1/2})^T$. In addition, let $C^{-1/2} = (C^{1/2})^{-1}$ and $C^{-T/2} = (C^{T/2})^{-1}$. Usual versions of the square root are the Cholesky square root and the symmetric positive definite square root. The left and right Cholesky square roots $C^{1/2}$ and $C^{T/2}$ are defined as the lower and upper triangular matrices with positive diagonal elements satisfying $C^{1/2}C^{T/2} = C$ and $C^{T/2} = (C^{1/2})^T$. Denote by $\| \, . \, \|_1$ the sum of the absolute values of the elements of a matrix. Also denote by $\lambda_{\min}(.)$ and $\lambda_{\max}(.)$ the minimum and the maximum eigenvalues of a symmetric matrix. For any fixed $A > 0$, define subsets of $\mathbb{R}^d$ by

$$N_n(A) = \{\boldsymbol{\theta} : (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbf{G}_n (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq A^2, \boldsymbol{\theta} \in \Theta\}.$$

To obtain the existence, consistency and the asymptotic distribution of an ME for the model, we need the following assumptions on the asymptotic behaviour of the first and second derivative matrices and their expectations. (Convergences are as $n \to \infty$ unless otherwise stated.)
(B1) $E\{\mathbf{S}_n(\boldsymbol{\theta}_0)\} = \mathbf{0}$, and the matrices $\mathbf{D}_n$ and $\mathbf{G}_n$ are finite, where the expectations are taken with respect to the true distributions.
(B2) $\lambda_{min}\{\mathbf{G}_n\} \to \infty$. (When (B2) holds, $\mathbf{G}_n$ is positive definite for n large enough, so we assume it to be so in general.)
(B3) $\sup_{\boldsymbol{\theta} \in N_n(A)} \| \mathbf{G}_n^{-1/2} \mathbf{F}_n(\boldsymbol{\theta}) \mathbf{G}_n^{-T/2} - \mathbf{I}_k \|_1 \xrightarrow{\text{P}} 0$.
(B4) For some positive definite matrix V, $\| \mathbf{G}_n^{-1/2} \mathbf{D}_n \mathbf{G}_n^{-T/2} - \mathbf{V} \|_1 \to 0$.
(B5) $\mathbf{D}_n^{-1/2} \mathbf{S}_n(\boldsymbol{\theta}_0) \to^D N(\mathbf{0}, \mathbf{I}_k)$.

Based on $E\left(\overline{X} - E\overline{X}\right) = \mathbf{0}$ we can conclude that $E\{\mathbf{S}_n(\boldsymbol{\theta})\} = \mathbf{0}$. So (B1) satisfies.

For (B2), we have $\lambda(\mathbf{G}_n) = n\lambda(\mathbf{G})$ and $\lambda > 0$ because $\mathbf{G}$ is positive definite.

$\mathbf{G}_n$ is a positive definite matrix (theorem 3.1) and nonsingular so $\mathbf{G}_n^{1/2}$ and $\mathbf{G}_n^{-1/2}$ does exist (Cholesky decomposition). Because $\mathbf{G}_n = \mathbf{D}_n$, we have $\mathbf{G}_n^{-1/2}\mathbf{D}_n\mathbf{G}_n^{-T/2} = \mathbf{I}$, so we can take $\mathbf{V} = \mathbf{I}$, a positive definite matrix. Therefore (B4) holds, too.

Also $\mathbf{D}_n$ is a nonsingular positive definite matrix (theorem 3.1), so, based on Cholesky decomposition theorem $\mathbf{D}_n^{-1/2}$ exist. We have
$E(\mathbf{D}_n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)) = \mathbf{D}_n^{-1/2}E(\mathbf{S}_n(\boldsymbol{\theta}_0)) = \mathbf{0}$ and
$\mathrm{Var}(\mathbf{D}_n^{-1/2}S_n(\boldsymbol{\theta}_0)) = \mathbf{D}_n^{-1/2}\mathrm{Var}(\mathbf{S}_n(\boldsymbol{\theta}_0))\mathbf{D}_n^{-T/2} = \mathbf{I}$. Also,

$$\mathbf{D}_n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0) = \frac{\mathbf{D}^{-1/2}\mathbf{M}\sum_{i=1}^{n}(\mathbf{X}_i - E\mathbf{X}_i)}{\sqrt{n}}, \tag{12}$$

thus from the Central Limit Theorem, $\mathbf{D}_n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$ is asymptotically standard normal as $n \to \infty$. So (B5) holds.

It remains to verify (B3). Fix $A > 0$, $n \geq 1$ and choose $\delta > 0$. Keep $(\kappa, \mu_1, \mu_2, \ldots, \mu_{d-1}) \in N_n(A)$, from the definition of $N_n(A)$,

$$\lambda_{\min}(\mathbf{G}_n) \leq \frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|}\mathbf{G}_n\frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|} \leq \frac{A^2}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2},$$

where $\lambda_{\min}(\mathbf{G}_n) = \inf_{|u|=1} u^T\mathbf{G}_n u$. On noting that $\lambda_{min}(\mathbf{G}_n) > 0$, we have

$$|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 \leq \frac{A^2}{n\lambda_{\min}(\mathbf{G})},$$

because it is clear that $\mathbf{G}_n = n\mathbf{G}$. So

$$|\kappa - \kappa_0| < \frac{A}{\sqrt{n\lambda_{\min}(\mathbf{G})}}, \tag{13}$$

and

$$|\mu_i - \mu_{i0}| < \frac{A}{\sqrt{n\lambda_{\min}(\mathbf{G})}} \quad \mathit{for} \quad i = 1, \ldots, d-1.$$

So we have $|\kappa| < |\kappa_0| + \frac{A}{\sqrt{n\lambda_{\min}(\mathbf{G})}}$ and $|\mu_i| < |\mu_{i0}| + \frac{A}{\sqrt{n\lambda_{\min}(\mathbf{G})}}$ for $i = 1, \ldots, d-1$. Also, we need to add another assumption that $|\mu_d| > \delta$, where $\delta > 0$.

To verify assumption (B3), it is sufficient to show that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in N_n(A)} \| n^{-1}\mathbf{F}_n(\boldsymbol{\theta}) - \mathbf{G}_0 \|_1 \\
&= \sup_{\boldsymbol{\theta}\in N_n(A)} \| n^{-1}\mathbf{F}_n(\boldsymbol{\theta}) - n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0) + n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0) - \mathbf{G}_0 \|_1 \\
&\leq \sup_{\boldsymbol{\theta}\in N_n(A)} \| n^{-1}\mathbf{F}_n(\boldsymbol{\theta}) - n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0) \|_1 + \| n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0) - \mathbf{G}_0 \|_1 \\
&\xrightarrow{\text{P}} 0, \qquad \text{as n} \to \infty.
\end{aligned}$$

where $\mathbf{G}_0 = n^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)$. From the weak law of large numbers, we have $n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0)\xrightarrow{\text{P}}\mathbf{G}_0$ as n goes to infinity. To show that $\| n^{-1}\mathbf{F}_n(\boldsymbol{\theta}) - n^{-1}\mathbf{F}_n(\boldsymbol{\theta}_0) \|_1 \xrightarrow{\text{P}} 0$, based on (13), we consider three elements separately. For the first element, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in N_n(A)} \left| \frac{-1}{\mu_d}\widetilde{\boldsymbol{\mu}}_i^T\overline{\mathbf{X}}_{i,d} + \frac{1}{\mu_{d0}}\widetilde{\boldsymbol{\mu}}_{i0}^T\overline{\mathbf{X}}_{i,d} \right| \\
&= \sup_{\boldsymbol{\theta}\in N_n(A)} \left| \frac{\widetilde{\boldsymbol{\mu}}_{i0}}{\mu_{d0}} - \frac{\widetilde{\boldsymbol{\mu}}_i}{\mu_d} \right|^T |\overline{\mathbf{X}}_{i,d}| \\
&= \sup_{\boldsymbol{\theta}\in N_n(A)} \left| \frac{\mu_{i0}}{\mu_{d0}} - \frac{\mu_i}{\mu_d} \right| |\overline{X}_d| \\
&\leq \sup_{\boldsymbol{\theta}\in N_n(A)} \left(\frac{|\mu_{i0}| - |\mu_i|}{|\mu_d|} + |\mu_{i0}|(\frac{1}{|\mu_{d0}|} - \frac{1}{|\mu_d|}) \right) |\overline{X}_d| \\
&\leq o(\sqrt{n})|\overline{X}_d|.
\end{aligned}$$

We have $|\overline{X}_d|\xrightarrow{\text{P}}E|X_d|$, from the weak law of large numbers when $n \to \infty$. Also $E|X_d|$ is bounded, so the last equation goes to zero as $n \to \infty$. If we use the same technique for the second and third elements in $\mathbf{F}_n(\boldsymbol{\theta})$, which respectively are $\left|\kappa\left(\frac{\mu_d^2+\mu_i^2}{\mu_d^3}\right)\overline{X}_d - \kappa_0\left(\frac{\mu_{d0}^2+\mu_{i0}^2}{\mu_{d0}^3}\right)\overline{X}_d\right|$ and $\left|\frac{\kappa}{\mu_d^3}\mu_i\mu_j\overline{X}_d - \frac{\kappa_0}{\mu_{d0}^3}\mu_{i0}\mu_{j0}\overline{X}_d\right|$, we can see that the sup of these on $N_n(A)$ tend to zero as $n \to \infty$.

We have now verified that all of the assumptions (B1)- (B5) hold for the von Mises distribution. Consider Ω to be a subset of Θ and satisfy assumption (A2) as below:
(A2) A subset Ω of Θ is said to satisfy (A2) if there is a closed cone C_Ω with vertex at $\boldsymbol{\theta}_0$ such that $C_\Omega \cap \mathcal{N} = \Omega \cap \mathcal{N}$, where $\mathcal{N}$ is a closed nonempty neighborhood in $\mathbb{R}^d$ of $\boldsymbol{\theta}_0$.
or a weaker condition
(A2$'$) A subset Ω of Θ is said to satisfy (A2$'$) if Ω contains $\boldsymbol{\theta}_0$, and if the intersection between Ω and a closed neighborhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$ is a closed subset of $\mathbb{R}^d$.

4 Hypothesis Testing

Our first aim is to test the simple hypothesis

$$\begin{cases} H_0 & : \kappa = \kappa_0, \mu_1 = \mu_{10}, \ldots, \mu_{d-1} = \mu_{d-10} \\ H_A & : Otherwise, \end{cases} \tag{14}$$

where $0 < \mu_{10} < 1$ and $\kappa_0 > 0$. So $\boldsymbol{\theta}_0$ is in Θ^0, the interior of the parameter space, $\boldsymbol{\theta}_0 \in \Theta^0$. To do this we find the distribution of the deviance statistic in (2)

Here Ω and τ are two fixed subsets of Θ which specify the subsets of the parameter space corresponding to the null and alternative hypotheses respectively.

They are required to satisfy the assumption (A2) with corresponding C_Ω and C_τ.

For the hypotheses specified in (14), we have $\Omega = \{(\kappa_0, \mu_{10}, \ldots, \mu_{d-10})\}$ and $\tau = \{(\kappa, \mu_1, \ldots, \mu_{d-1}) \in (0, \infty) \times (0,1) \times \ldots, \times (0,1)$
$- (\kappa_0, \mu_{10}, \ldots, \mu_{d-10})\}$.

Let $T_n = \frac{1}{\sqrt{n}} \mathbf{I}_{d\times d}$ and define

$$\widetilde{C}_{\Omega_n} = \{\widetilde{\boldsymbol{\theta}} : \widetilde{\boldsymbol{\theta}} = T_n \mathbf{G}_n^{T/2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \boldsymbol{\theta} \in C_\Omega\} \tag{15}$$

and similarly for $\widetilde{C}_{\tau_n}$. Then

$$\widetilde{C}_{\Omega_n} = \{(0, 0, \ldots, 0)\} \quad \text{and} \quad \widetilde{C}_{\tau_n} = \mathbb{R}^d. \tag{16}$$

Now we can check the following assumption, (A3), of Vu & Zhou (1997). (A3) is satisfied if there exists a closed cone $\widetilde{C}_\Omega$ with vertex at 0, not depending on n, such that the sets $\widetilde{C}_{\Omega_n}$ asymptotically coincide with $\widetilde{C}_\Omega$ in the sense that as $n \to \infty$,

$$\sup_{|\beta|=1} \left| \inf_{\boldsymbol{\theta} \in \widetilde{C}_{\Omega_n}} |\beta - \boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} |\beta - \boldsymbol{\theta}|^2 \right| \to 0.$$

On noting that $\widetilde{C}_{\Omega_n}$ and $\widetilde{C}_{\tau_n}$ as defined in (16) are not dependent on n, we can take

$$\widetilde{C}_\Omega = \{(0, 0, \ldots, 0)\} \quad \text{and} \quad \widetilde{C}_\tau = \mathbb{R}^d. \tag{17}$$

Then (A3) holds.

Lemma 4.1. *The asymptotic distribution of d_n for (14) is a chi square distribution with d degrees of freedom.*

Proof. Based on Theorem 2.2 in Vu & Zhou (1997), because $\mathcal{L}_n$ and its first and second derivatives exist and are continuous functions on $\boldsymbol{\theta} a \cap \mathcal{N}$, (A3) holds and also (B1) - (B5) hold. So the asymptotic distribution of d_n exists and is the same as the distribution of

$$\inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} ||\mathbf{N} - \boldsymbol{\theta}||^2 - \inf_{\boldsymbol{\theta} \in \widetilde{C}_\tau} ||\mathbf{N} - \boldsymbol{\theta}||^2, \tag{18}$$

where $\mathbf{N} = (N_1, N_2, \ldots, N_d)^T$ is a random vector which has a multivariate normal distribution with mean zero and identity matrix I and $\boldsymbol{\theta}$ is a d - dimensional vector.
Based on (17), we have

$$\inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} ||\mathbf{N} - \boldsymbol{\theta}||^2 = N_1^2 + N_2^2 + \ldots + N_d^2.$$

While the second inf is over $C_\tau = \mathbb{R}^d$, so it equals zero. Therefore we can conclude that the distribution of d_n is the same as the asymptotic distribution of $N_1^2 + N_2^2 + \ldots + N_d^2$, the sum of d standard Normal variables, that is a chi square with d degrees of freedom. □

Lemma 4.2. *d_n for the hypothesis test*

$$\begin{cases} H_0 & : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ H_A & : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \end{cases} \tag{19}$$

has a chi square distribution with $d-1$ degrees of freedom.

Proof. In this simple hypothesis, Ω is $(0,\infty) \times \{\boldsymbol{\mu}_0\}$, $\mathcal{N}$ is a neighborhood around $\kappa \times \boldsymbol{\mu}_0$ in $\mathbb{R}$, C_Ω becomes $x_d - axis$, and $\widetilde{C}_\Omega$ is $\mathbb{R}$. Therefore, the distance of $N = (N_1, N_2, \dots, N_d)^T$ from $\widetilde{C}_\Omega$ is $\sqrt{N_1^2 + N_2^2 + \dots + N_{d-1}^2}$. For the alternative hypothesis, τ is $(0,\infty) \times \{\{(-1,+1) \times (-1,+1) \times \dots \times (-1,+1)\} - \boldsymbol{\mu}_0\}$, the centre of C_τ is at $\kappa, \boldsymbol{\mu}$, and $\mathcal{N}$ is a ball around $\kappa, \boldsymbol{\mu}$. Therefore, $\widetilde{C}_\tau$ is $\mathbb{R}^d$. The distance of $N = (N_1, N_2, \dots, N_d)$, which is in $\mathbb{R}^d$, from $\mathbb{R}^d$ is zero. Thus

$$\inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} ||\mathbf{N} - \boldsymbol{\theta}||^2 - \inf_{\boldsymbol{\theta} \in \widetilde{C}_\tau} ||\mathbf{N} - \boldsymbol{\theta}||^2 = N_1^2 + N_2^2 + \dots + N_{d-1}^2,$$

which has a chi square distribution with $d-1$ degrees of freedom. □

Lemma 4.3. *For the hypothesis test*

$$\begin{cases} H_0 & : \kappa = \kappa_0 \\ H_A & : \kappa \neq \kappa_0, \end{cases} \tag{20}$$

d_n has a chi square distribution with one degree of freedom.

Proof. $\Omega = \kappa_0 \times (-1,+1) \times \dots \times (-1,+1)$, $\mathcal{N}$ is a neighborhood around κ_0, C_Ω is a hyperplane which goes through κ_0, and finally $\widetilde{C}_\Omega$ becomes $(x_2, x_3, \dots, x_d)$ hyperplane. As a result $\inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} ||\mathbf{N} - \boldsymbol{\theta}|| = N_1^2$, $\tau = \{[0,\infty) - \kappa_0\} \times (-1,+1) \times \dots \times (-1,+1)$, $\mathcal{N}$ is a ball centred at (κ_0, μ), $\mu \in (-1,+1) \times \dots \times (-1,+1)$, C_τ is a d dimensional plane in $\mathbb{R}^d$, and $\widetilde{C}_\tau$ is $\mathbb{R}^d$ centred at $\mathbf{0}_d$. We have

$$\inf_{\boldsymbol{\theta} \in \widetilde{C}_\Omega} ||\mathbf{N} - \boldsymbol{\theta}|| - \inf_{\boldsymbol{\theta} \in \widetilde{C}_\tau} ||\mathbf{N} - \boldsymbol{\theta}|| = N_1^2 - 0 = N_1^2 \tag{21}$$

and has a chi square distribution with one degree of freedom. □

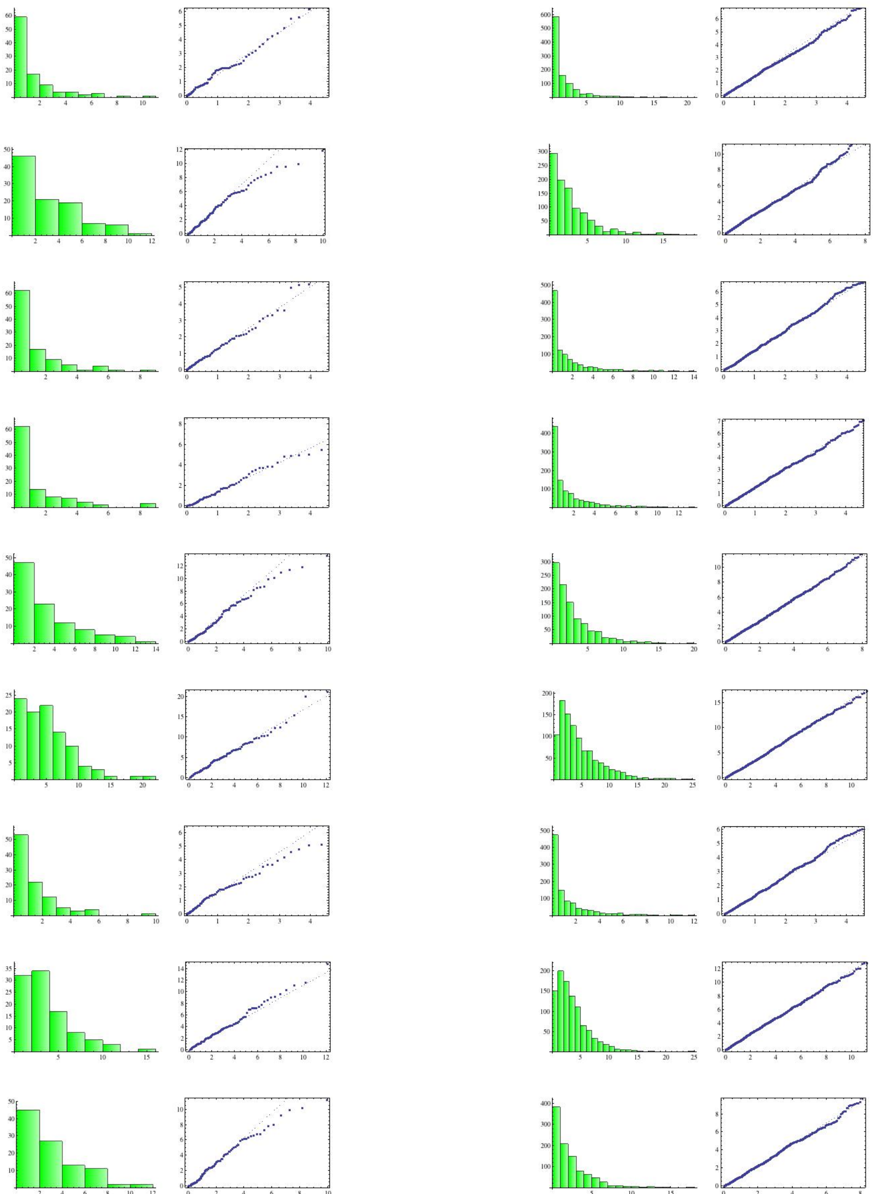

Figure 1: Histograms and Quantile plots of d_n for different tests in Table 1; Right is for $n = 100$ and Left is when $n = 1000$; Each row in this Figure coincides with the number of row in the Table 1

Table 1: Simulation result for the hypothesis tests of the parameters of a von Mises Fisher distribution in different dimensions; the number in the columns of Figure reflects, relatively, the number of row and column in Figure 1

Row	Dimension	n	r	H_0	Dis. of d_n	Est. df	Figure
1	2	100	100	$\kappa = 10$	$\chi^2(1)$	1.28	1,1
	2	100	1000	$\kappa = 10$	$\chi^2(1)$	1.19	1,2
2	2	100	100	$\kappa = 10, \mu_1 = 0.5$	$\chi^2(2)$	2.51	2,1
	2	100	1000	$\kappa = 10, \mu_1 = 0.5$	$\chi^2(2)$	2.49	2,2
3	2	100	100	$\mu_1 = 0.5$	$\chi^2(1)$	1.17	3,1
	2	100	1000	$\mu_1 = 0.5$	$\chi^2(1)$	1.11	3,2
4	3	100	100	$\kappa = 10$	$\chi^2(1)$	1.04	4,1
	3	100	1000	$\kappa = 10$	$\chi^2(1)$	1.18	4,2
5	3	100	100	$\kappa = 10, \mu_1 = 0.5$	$\chi^2(2)$	2.68	5,1
	3	100	1000	$\kappa = 10, \mu_1 = 0.5$	$\chi^2(2)$	2.43	5,2
6	3	100	100	$\kappa = 10, \mu_1 = 0.5, \mu_2 = 0.1$	$\chi^2(3)$	4.61	6,1
	3	100	1000	$\kappa = 10, \mu_1 = 0.5, \mu_2 = 0.1$	$\chi^2(3)$	4.09	6,2
7	10	100	100	$\kappa = 10$	$\chi^2(1)$	1.10	7,1
	10	100	1000	$\kappa = 10$	$\chi^2(1)$	1.27	7,2
8	10	100	100	$\mu_1 = \mu_2 = \mu_3 = 0$	$\chi^2(3)$	3.49	8,1
	10	100	1000	$\mu_1 = \mu_2 = \mu_3 = 0$	$\chi^2(3)$	3.38	8,2
9	10	100	100	$\mu_1 = \mu_2 = \mu_3$	$\chi^2(2)$	2.27	9,1
	10	100	1000	$\mu_1 = \mu_2 = \mu_3$	$\chi^2(2)$	2.10	9,2

5 Simulation results

Row one in Table 1 and Figure 1 show the results of simulations from a 2 dimensional von Mises Fisher distribution when we test $\kappa = 10$ and calculate d_n for this test. In order to do this, we simulate $n = 100$ data from a 2-dimensional von Mises distribution with $\kappa = 10$ and calculate the value of d_n for the test

$$\begin{cases} H_0 & : \kappa = 10 \\ H_A & : \kappa \neq 10 \end{cases}$$

based on the formula of d_n in (2). Table 1 shows these results, while considering two different replications of $r = 100$ and $r = 1000$ to calculate d_n. The program is written in Mathematica and "Est. df" in Table 1 is the estimated degree of freedom in a chi square distribution and is the mean of the data.

The column H_0 in Table 1 describes the null hypothesis which we test. The alternative hypotheses are in the form of non-equalities for the rows 1 to 7 and for the columns 8 and 9 the alternatives are "at least one equality is not satisfied". The distribution of d_n is chi square with the degrees of freedom calculated based on the methodology in Lemmas 4.1 to 4.3. There are totally 18 figures in 1 that are in 9 rows and 2 columns. The number of rows and columns are written in the last column of Table 1. For example, the number 1,2 shows the figure which is in the first row and second column of Figure 1.

Acknowledgements
I deeply appreciate my supervisor, Professor Ross Maller from the Australian National University, who devoted his knowledge and time to support me even after the graduation.

References

[1] D. Chant, On asymptotic tests of composite hypotheses in nonstandard conditions, *Biometrika*, 61 (1974).

[2] H. Chernoff, On the distribution of the likelihood ratio, *The Annals of Mathematical Statistics*, 25 (1954).

[3] P. I. Feder, On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions, *The Annals of Mathematical Statistics*, 39 (1968).

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

[5] C. J. Geyer, On the asymptotics of constrained M-estimation, *The Annals of Statistics*, 22 (1994).

[6] M. Ghodsi, Nonstandard Estimation for the von Mises Fisher Distribution, *Bulletin of the Australian Mathematical Society*, PhD Thesis, First published online: 07 March 2018.

[7] K. V. Mardia and P. E. Jupp, *Directional statistics*, Wiley, John Wiley & Sons, Inc. (2000).

[8] P. A. P. Moran, Maximum-likelihood estimation in non-standard conditions, *Mathematical Proceedings of the Cambridge Philosophical Society*, 70 (1971).

[9] S. G. Self and K.-Y. Liang, Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association*, 82 (1987).

[10] M. J. Silvapulle and P. K. Sen, *Constrained statistical inference: order, inequality, and shape constraints*, Wiley, Hoboken, NJ (2011).

[11] H. T. V. Vu and S. Zhou, Generalization of likelihood ratio tests under nonstandard conditions, *The Annals of Statistics*, 25 (1997).

[12] G. S. Watson, *Statistics on spheres*, University of Arkanas Lecture Notes in the Mathematical Sciences 6, Wiley, New York (1983).

[13] S. S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, *The Annals of Mathematical Statistics*, 9 (1938).

Maryam Ghodsi
Department of Mathematics and Statistics
Lecturer
Jahrom Branch, Islamic Azad University
Jahrom, Iran
E-mail: ghodsi.maryam@gmail.com