Testing homogeneity in a mixture of von Mises distributions with a structural parameter

Jiahua CHEN, Pengfei LI and Yuejiao FU

Key words and phrases: Asymptotic distribution; circular data; directional data; mixture of von Mises distributions; modified likelihood; penalty.
MSC 2000: Primary 62F03; secondary 62F05.

Abstract: The modified likelihood ratio test has been successfully applied for the homogeneity test in a variety of mixture models. In this paper, the authors propose the use of the modified likelihood ratio test and the iterative modified likelihood ratio test in general two-component von Mises mixture with a structural parameter. Two accuracy enhancing methods are developed. The limiting distributions of the resulting test statistics are derived. Simulations show that the test statistics have accurate type I errors and adequate power. A data set on orientations of turtles is analyzed, and the result suggests that there exist two subgroups of turtles that traveled in opposite directions.

Title in French: we can supply this

Résumé : Insérer votre résumé ici. The modified likelihood ratio test has been successfully applied for the homogeneity test in a variety of mixture models. In this paper, the authors propose the use of the modified likelihood ratio test and the iterative modified likelihood ratio test in general two-component von Mises mixture with a structural parameter. Two accuracy enhancing methods are developed. The limiting distributions of the resulting test statistics are derived. Simulations show that the test statistics have accurate type I errors and adequate power. A data set on orientations of turtles is analyzed, and the result suggests that there exist two subgroups of turtles that traveled in opposite directions.

1. INTRODUCTION

Circular data, which are measured in the form of angles or two dimensional orientations, arise commonly in many disciplines, including astronomy, biology, ecology, geology, physics and medicine. Examples of such data include direction of flight of birds or the orientation of animals, wind and ocean current directions, circadian and other biorhythms. Describing and analyzing such data statistically poses a lot of interesting and challenging problems. For example, the sample mean defined for linear data is no longer appropriate to measure the center of circular data. Suppose two turtles moved at 10° and 350° measured clockwise from north. Their arithmetic mean is 180°, due south, while the two movements point toward north. Some other special features of circular distributions will be given in Sections 2 and 5.

(2001) presents the latest developments in circular statistics. Statistical inferences in circular or directional mixture models have been discussed by many authors, such as Stephens (1969), Fraser, Hsu & Walker (1981), Hsu, Walker & Ogren (1986), Kim & Koo (2000), and Holzmann, Munk & Stratmann (2004).

As a circular analog of the normal distribution on the real line, the von Mises distribution is the most commonly used distribution for circular data. See the aforementioned references for general properties of the von Mises distribution. When the data are drawn from a heterogeneous population, mixtures of von Mises distributions are often used, see Grimshaw, Whiting & Morris (2001) for an example in geology. Before one tries to fit a mixture model, it might be of value to know whether the data arise from a homogeneous or heterogeneous population. If the data are homogeneous, it is not even necessary to go into mixture modeling. Similar to the linear case, unless the components are fairly distinct, testing the order or the number of components in a circular mixture is a challenging problem.

The following is an illustrating example involving movements of turtles due to Gould (1959), which was first analyzed by Stephens (1969). The orientation of a certain type of turtle was studied through a designed experiment. 76 turtles were released from an experimental site. Their initial travel directions were measured clockwise from north. The data shows that the majority of the turtles traveled in directions roughly North-East (‘homeward’), while some turtles went in almost the opposite direction and a few went in other directions. One interpretation is that the turtles have a preferred direction (‘homeward’), but some confuse forwards with backwards (Stephens, 1969 pp. 20). The statistical problem of interest here is to test whether there is a group of turtles that confuse forwards with backwards.

Fu, Chen & Li (2007) discuss the modified likelihood ratio test (MLRT) for homogeneity in a mixture of von Mises distributions with applications in biological circadian clock studies. It is the first work using the modified likelihood approach in finite mixture models for circular data analysis and the mixture model considered has a special feature that one mean direction in the mixture is known. The focus of the current paper is on testing homogeneity in a more general two-component von Mises mixture, where the two mean directions are both unknown. The technical components under the new model pose a greater challenge. In addition, the approximation of the limiting distribution of the MLRT statistic becomes less satisfactory. We suggest two ways to improve the approximation. Furthermore, a novel iterative procedure proposed by Li & Chen (2007) is employed to this problem. It not only retains the nice asymptotic properties of the MLRT but also enjoys superior finite sample performance.

The paper is organized as follows. In Section 2, we study the circular moment property for a mixture of two von Mises distributions. The results suggest that the structural parameter tends to be overestimated when a two-component von Mises mixture is used to fit the data arising from a homogeneous von Mises distribution. This phenomenon motivates an additional penalty on the large values of the structural parameter. The main results are developed in Section 3 and the proofs are deferred to the Appendix. In Section 4, we provide some simulation results and one real data example. The conclusions are given in Section 5.

2. THE VON MISES DISTRIBUTION AND CIRCULAR MOMENTS

The von Mises distribution was first introduced as a statistical model by von Mises (1918). It plays a key role in statistical inference on the unit circle. Because of its importance and similarities to the Normal distribution on the line, it is also called the Circular Normal distribution. The von Mises distribution $M(\mu, \kappa)$ has probability density function (pdf)

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad |\theta| \leq \pi,$$

where $|\mu| \leq \pi$ and $\kappa \geq 0$. The function $I_0(\kappa)$ is the normalizing constant.
The special feature of circular data is directional and the measures of location and dispersion are defined differently from those for linear data. Let \( \theta \) be a circular random variable. The circular mean direction and the circular variance can be defined as

\[
\text{CE}(\theta) = \arg\min_{\mu \in [-\pi, \pi]} E\{2 - 2\cos(\theta - \mu)\}
\]

and

\[
\text{CVar}(\theta) = E[2 - 2\cos(\theta - \text{CE}(\theta))],
\]

respectively. We may note that \( 2 - 2\cos(\theta - \mu) = 4\sin^2((\theta - \mu)/2) \). Hence, the circular mean is the minimum point of a trigonometry distance, and the circular variance is the resulting minimum value. Replacing the sine function in the definition by the identity function leads to the usual mean and variance for linear data.

For the von Mises distribution \( M(\mu, \kappa) \), the mean direction is \( \mu \) and the circular variance is \( 2 - 2A(\kappa) \), where

\[
A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}
\]

and

\[
I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) \exp(\kappa \cos \theta) \, d\theta.
\]

As \( \kappa \) increases, the circular variance decreases and the distribution places more mass around the mean direction. Hence \( \kappa \) is also called the concentration parameter.

We now give the circular mean and variance for a mixture of two von Mises distributions. Let \( \theta \) be a circular random variable with distribution \( (1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa) \) for some \( \kappa > 0 \). Then, for any \( \mu \), we have

\[
E\{\cos(\theta - \mu)\} = A(\kappa) \cos(\eta - \mu) \sqrt{1 - 4\alpha(1 - \alpha) \sin^2(\frac{\mu_1 - \mu_2}{2})},
\]

where \( \eta \in [-\pi, \pi] \) is an angle such that

\[
\cos \eta = \{(1 - \alpha) \cos \mu_1 + \alpha \cos \mu_2\}/\sqrt{1 - 4\alpha(1 - \alpha) \sin^2(\frac{\mu_1 - \mu_2}{2})}
\]

and

\[
\sin \eta = \{(1 - \alpha) \sin \mu_1 + \alpha \sin \mu_2\}/\sqrt{1 - 4\alpha(1 - \alpha) \sin^2(\frac{\mu_1 - \mu_2}{2})}.
\]

Hence, we have

\[
\text{CE}(\theta) = \arg\max_{\mu} E\{\cos(\theta - \mu)\} = \eta.
\]

Consequently, by (1)

\[
\text{CVar}(\theta) = 2 - 2A(\kappa) \sqrt{1 - 4\alpha(1 - \alpha) \sin^2(\frac{\mu_1 - \mu_2}{2})}.
\]

It is seen that the variance of a heterogeneous model, where \( \alpha(1 - \alpha) \neq 0 \) and \( \mu_1 \neq \mu_2 \), is larger than that of a homogeneous model with the same \( \kappa \). Thus, because \( A(\kappa) \) is an increasing function of \( \kappa \), fitting a heterogeneous model to data arising from a homogeneous model tends to result in a larger fitted concentration parameter.

3. MAIN RESULTS

Assume that a circular random sample \( \theta_1, \ldots, \theta_n \) is drawn from the von Mises mixture distribution \( (1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa) \), where \( 0 \leq \alpha \leq 1, -\pi \leq \mu_1, \mu_2 \leq \pi \) and \( \kappa \geq 0 \). We are interested in testing

\[
H_0 : \alpha(1 - \alpha)(\mu_1 - \mu_2) = 0
\]
versus the full model. This section focuses on the asymptotic properties of likelihood-based testing procedures.

3.1. The likelihood ratio test.

The log-likelihood function can be expressed as

\[ l_n(\alpha, \mu_1, \mu_2, \kappa) = -n \log I_0(\kappa) + \sum \log[(1 - \alpha) \exp\{\kappa \cos(\theta_i - \mu_1)\} + \alpha \exp\{\kappa \cos(\theta_i - \mu_2)\}] \]

Let \( \hat{\mu}_0 \) and \( \hat{\kappa}_0 \) be the MLEs under the null hypothesis and let \( \hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\kappa} \) be the MLEs under the full model. The likelihood ratio test (LRT) statistic is defined by

\[ R_n = 2\{l_n(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}) - l_n(0.5, \hat{\mu}_0, \hat{\kappa}_0)\} \]

Without loss of generality, let \( M(0, \kappa_0) \) be the null distribution.

The MLEs have some interesting properties as stated in the following two lemmas. Our proofs, which are omitted, are similar to arguments in Fu, Chen & Li (2007).

**Lemma 1.** Assume that the distribution of the random sample \( \theta_1, \ldots, \theta_n \) is given by \( M(0, \kappa_0) \) for some \( \kappa_0 > 0 \). Let \( \hat{\kappa} \) be the MLE of \( \kappa \) under the full model \( (1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa) \). Then there exists a constant \( 0 < \Delta < \infty \) such that \( \lim_{n \to \infty} P(\hat{\kappa} \leq \Delta) = 1 \).

As a consequence of Lemma 1, the parameter space under consideration can be reduced to a compact one for theoretical derivations. With identifiability (Fraser, Hsu & Walker 1981, Holzmann, Munk & Stratmann 2004), Lemma 1 implies the partial consistency of the MLEs.

**Lemma 2.** Assume that the distribution of the random sample \( \theta_1, \ldots, \theta_n \) is given by \( M(0, \kappa_0) \). Let \( \hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\kappa} \) be the MLEs of \( \alpha, \mu_1, \mu_2, \) and \( \kappa \) under the full model \( (1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa) \). Then \( (1 - \hat{\alpha})\hat{\mu}_1 + \alpha \hat{\mu}_2 \to 0, (1 - \hat{\alpha})\hat{\mu}_1^2 + \alpha \hat{\mu}_2^2 \to 0 \) and \( \hat{\kappa} \to \kappa_0 \) in probability, as \( n \to \infty \).

The asymptotic distribution of the LRT statistic is given in the following theorem and the proof is deferred to the Appendix. Let \( \xrightarrow{d} \) denote convergence in distribution.

**Theorem 1.** Let \( \theta_1, \ldots, \theta_n \) be a random sample from the mixture distribution \( (1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa) \), where \( 0 \leq \alpha \leq 1, -\pi \leq \mu_1, \mu_2 \leq \pi \) and \( \kappa \geq 0 \). Let \( R_n \) be the LRT statistic for testing \( H_0 : \alpha(1 - \alpha)(\mu_1 - \mu_2) = 0 \). Then under the null distribution \( M(0, \kappa_0) \), as \( n \to \infty \),

\[ R_n \xrightarrow{d} \sup_{|\mu| \leq \pi} \{\zeta^+(\mu)\}^2, \]

where \( \zeta(\mu), |\mu| \leq \pi, \) is a Gaussian process with mean 0, variance 1 and autocorrelation \( \rho(s, t) \) which is given by

\[ \rho(s, t) = \frac{g(s, t)}{(g(s, s)g(t, t))^{\frac{1}{2}}}, \quad \text{for } s, t \neq 0, \]

where

\[ g(s, t) = \frac{1}{st} \left[ \frac{I_0(\kappa_0)[\{(\cos s + \cos t) - 1\}^2 + (\sin s \sin t)^2]^{\frac{1}{2}}}{I_0(\kappa_0)} - \right. \]

\[ \left. \frac{A^2(\kappa_0)(\cos s - 1)(\cos t - 1)}{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)} - \kappa_0 A(\kappa_0) \sin s \sin t \right]. \]
Chen & Chen (2003) gives the asymptotic distribution of the LRT statistic in a two-component normal mixture with a structural parameter. In comparison with the result above, their Gaussian process contains a spike at 0. More discussions on the subtle difference between normal and von Mises mixtures will be given in Section 5.

3.2. The modified likelihood approaches.

The asymptotic distribution of the LRT statistic is not convenient to use in practice, see the discussions in Chen & Chen (2001, 2003), Dacunha-Castelle & Gassiat (1999), and Liu & Shao (2003). The MLRT proposed in Chen (1998) and Chen, Chen & Kalbfleisch (2001, 2004), provides a nice remedy to the ordinary LRT. We define the modified likelihood as

\[ pl_n(\alpha, \mu_1, \mu_2, \kappa) = l_n(\alpha, \mu_1, \mu_2, \kappa) + p(\alpha), \]  

(2)

where \( p(\alpha) \) is a continuous function such that \( p(\alpha) \to -\infty \) as \( \alpha \to 0 \) or 1 and it attains its maximal value at \( \alpha = 0 \).

The result in Theorem 1 is instrumental in analyzing the asymptotic properties of the MLRT. The limiting distribution of the MLRT is found to be a mixture of chi-squared distributions, which is very convenient to implement in practice.

Simulation studies indicate that the finite sample distribution of the above MLRT statistic under the null model is not well approximated by the null limiting distribution unless the sample size is very large. In the next subsection, we discuss two accuracy enhancing methods.

3.2.1. Accuracy enhancing methods.

Due to the moment properties discussed in Section 2, \( \kappa \) tends to be overestimated by the MLE or the modified MLE under the heterogeneous model. As a consequence, the finite sample distribution of the MLRT statistic is stochastically larger than the limiting distribution which inflates the type I error rate. To overcome this problem, we propose penalizing the fit with larger values of \( \kappa \). In particular, we suggest adding \( -\log(\kappa + 1) \) to \( pl_n \) in (2) and the resulting modified likelihood function becomes

\[ pl_n(\alpha, \mu_1, \mu_2, \kappa) = l_n(\alpha, \mu_1, \mu_2, \kappa) - \log(\kappa + 1) + p(\alpha). \] 

(3)

The corresponding MLRT statistic \( M_n^* \) is defined in the same fashion as \( M_n^* \). The second enhancing method is to select a more effective \( p(\alpha) \). The common choice for \( p(\alpha) \) is \( p(\alpha) = C \log\{4\alpha(1-\alpha)\} \), see Chen, Chen & Kalbfleisch (2001) and Zhu & Zhang (2004). We choose \( p(\alpha) = C \log(1 - |1 - 2\alpha|) \), as suggested by Li & Chen (2007). Both enhancing methods are found to be effective as demonstrated by simulation studies.

3.2.2. The iterative modified likelihood ratio test.

The iterative modified likelihood ratio test (IMLRT) is first proposed by Li & Chen (2007) for mixture models with single component parameter. Compared to the MLRT, it has wider applicability and often has better finite sample performance. In this subsection, we show that the IMLRT can be adapted to the problem with a structural parameter. Under the current model, we introduce the IMLRT procedure as follows.
Step 0: Let \( j = 1 \) and \( k = 0 \). Choose two integers \( K \geq 0 \), \( J > 0 \) and a number of initial \( \alpha \) values, say \( 0 < \alpha_1, \alpha_2, \ldots, \alpha_J \leq 0.5 \). Compute \( (\mu_0^*, \kappa_0^*) = \arg\max_{\mu, \kappa} \log \{p_n(1/2, \mu, \kappa)\} \) with \( p_n \) given by (3).

Step 1: Let \( \alpha_j^{(k)} = \alpha_j \).

Step 2: Compute

\[
(\mu_{j1}^{(k)}, \mu_{j2}^{(k)}, \kappa_j^{(k)}) = \arg\max_{\mu_1, \mu_2, \kappa} \log \{p_n(\alpha_j^{(k)}, \mu_1, \mu_2, \kappa)\}
\]

and

\[
M_n(\alpha_j^{(k)}) = 2\{p_n(\alpha_j^{(k)}, \mu_{j1}^{(k)}, \mu_{j2}^{(k)}, \kappa_j^{(k)}) - p_n(1/2, \mu_0^*, \kappa_0^*)\}.
\]

Step 3: For \( i = 1, 2, \ldots, n \), compute

\[
w_{ki} = \frac{\alpha_j^{(k)} f(X_i; \mu_{j1}^{(k)}, \kappa_j^{(k)})}{(1 - \alpha_j^{(k)}) f(X_i; \mu_{j1}^{(k)}, \kappa_j^{(k)}) + \alpha_j^{(k)} f(X_i; \mu_{j2}^{(k)}, \kappa_j^{(k)})}.
\]

Let

\[
\alpha_j^{(k+1)} = \arg\max_{\alpha} \left\{ (n - \sum_{i=1}^n w_{ki}) \log(1 - \alpha) + \sum_{i=1}^n w_{ki} \log(\alpha) \right\},
\]

\[
\mu_{j1}^{(k+1)} = \arg\max_{\mu_1} \left\{ \sum_{i=1}^n (1 - w_{ki}) \cos(\theta_i - \mu_1) \right\},
\]

\[
\mu_{j2}^{(k+1)} = \arg\max_{\mu_2} \left\{ \sum_{i=1}^n w_{ki} \cos(\theta_i - \mu_2) \right\},
\]

\[
\kappa_j^{(k+1)} = \arg\max_{\kappa} \left[ \kappa \left( \sum_{i=1}^n (1 - w_{ki}) \cos(\theta_i - \mu_{j1}^{(k+1)}) + \sum_{i=1}^n w_{ki} \cos(\theta_i - \mu_{j2}^{(k+1)}) \right) - n \log(I_0(\kappa)) - \log(\kappa + 1) \right].
\]

Compute

\[
M_n(\alpha_j^{(k+1)}) = 2\{p_n(\alpha_j^{(k+1)}, \mu_{j1}^{(k+1)}, \mu_{j2}^{(k+1)}, \kappa_j^{(k+1)}) - p_n(1/2, \mu_0^*, \kappa_0^*)\}.
\]

Let \( k = k + 1 \) and repeat Step 3 until \( k > K \).

Step 4: Let \( j = j + 1 \), \( k = 0 \) and go to Step 1, until \( j > J \).

Step 5: Calculate the IMLRT statistic

\[
IM_n^{(K)} = \max\{M_n(\alpha_j^{(K)}) \mid j = 1, 2, \ldots, J\}.
\]

Theorem 2. Let \( \theta_1, \ldots, \theta_n \) be a random sample from the von Mises mixture distribution \((1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa)\), where \( 0 < \alpha < 1, -\pi \leq \mu_1, \mu_2 \leq \pi, \kappa \geq 0 \). Suppose that \( p(\alpha) \) is a continuous function such that \( p(\alpha) \to -\infty \) as \( \alpha \to 0 \) or 1 and it attains its maximal value at \( \alpha = 0.5 \). Then under the null distribution \( M(0, \kappa_0) \), as \( n \to \infty \),

(a) \( M_n \overset{d}{\to} \chi^2_0 + \frac{1}{2} \chi^2_1 \);

(b) if one of the initial \( \alpha \) values is equal to 0.5, then for any nonnegative and finite integer \( K \),

\[
IM_n^{(K)} \overset{d}{\to} \chi^2_0 + \frac{1}{2} \chi^2_1,
\]

where \( \frac{1}{2} \chi^2_0 + \frac{1}{2} \chi^2_1 \) denotes an equal mixture of a distribution with point mass at zero and a \( \chi^2_1 \) distribution.
Intuitively, we should choose large values of $J$ and $K$ to ensure the efficiency of the test. However, our simulation suggests that $IM_n^{(1)}$ with $(\alpha_1, \alpha_2, \alpha_3) = (0.1, 0.3, 0.5)$ captures most power of $M_n$, which was also observed in Li & Chen (2007).

4. SIMULATIONS AND APPLICATION

Simulation studies were conducted to assess the performance of the proposed testing procedures. Let $R_n^* = 2\sup_{\mu_1, \mu_2, \kappa} l_n(1/2, \mu_1, \mu_2, \kappa) - l_n(1/2, \tilde{\mu}_0, \tilde{\kappa}_0)$. Let $IM_n^{(K)}$, $M_n$ and $M_n^*$ denote the IMLRT, the MLRT with $pl_n$ defined by (3) and the MLRT with $pl_n$ defined by (2), respectively.

The penalty $p(\alpha)$ for all those tests is chosen to be $p(\alpha) = C\log(1 - |1 - 2\alpha|)$. The choice of $C$ has some influences on the type I error of the test. Simulation studies are often used to find a suitable range of $C$. In the current testing problem, $C = 3$ has been found satisfactory.

The empirical null distributions of the test statistics were calculated based on 10000 replications for various combination of $n(= 100, 200)$ and $\kappa(= 2, 3)$. The simulated null rejection rates of the above test statistics are reported in Table 1. Clearly, the simulated type I error rates of $M_n^*$ are much larger than nominal values and the problem gets worse with larger $\kappa$. To save space, we omitted the results for $\kappa = 5$, but the trend is the same. Note that increasing $\kappa$ should lower the simulated levels. However, even if $C = \infty$, at which $M_n^*$ reduces to $R_n^*$, the simulated type I errors are still larger than the nominal levels. On the other hand, the simulated null rejection rates of $M_n$ and $IM_n^{(K)}$ are very close to the nominal levels and they do not depend much on the sample size nor on the size of the concentration parameter. Thus the penalty function $-\log(\kappa + 1)$ is very important.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Level</th>
<th>$IM_n^{(0)}$</th>
<th>$IM_n^{(1)}$</th>
<th>$M_n$</th>
<th>$M_n^*$</th>
<th>$R_n^*$</th>
<th>$IM_n^{(0)}$</th>
<th>$IM_n^{(1)}$</th>
<th>$M_n$</th>
<th>$M_n^*$</th>
<th>$R_n^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10</td>
<td>10.3</td>
<td>10.4</td>
<td>10.7</td>
<td>13.3</td>
<td>12.0</td>
<td>10.1</td>
<td>10.3</td>
<td>10.7</td>
<td>15.0</td>
<td>13.6</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>5.1</td>
<td>5.3</td>
<td>5.4</td>
<td>7.2</td>
<td>6.3</td>
<td>5.0</td>
<td>5.1</td>
<td>5.5</td>
<td>8.6</td>
<td>7.0</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1.0</td>
<td>1.2</td>
<td>1.2</td>
<td>1.8</td>
<td>1.4</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>2.2</td>
<td>1.6</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>9.8</td>
<td>9.8</td>
<td>10.0</td>
<td>11.9</td>
<td>11.2</td>
<td>10.0</td>
<td>10.0</td>
<td>13.5</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>5.1</td>
<td>5.1</td>
<td>5.3</td>
<td>6.7</td>
<td>5.9</td>
<td>5.1</td>
<td>5.1</td>
<td>5.4</td>
<td>7.6</td>
<td>6.5</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>1.5</td>
<td>1.2</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

To compare the power of the tests, we considered the following alternative models

$$(1 - \alpha)M(\pi/2, \kappa) + \alpha M(-\pi/2, \kappa),$$

with $\alpha = 0.05, 0.10, 0.25$ and $\kappa = 2, 3$. Simulated critical values were used for power calculation and the power was calculated based on 10000 repetitions. The results are in Table 2. It is seen that with one iteration and three initial values for $\alpha$, the test based on $IM_n^{(1)}$ captures most power of $M_n$.

We now apply the MLRT and the IMLRT with von-Mises kernel to the turtle data discussed earlier. The modified MLEs are $\tilde{\alpha}^* = 0.220$, $\tilde{\mu}_1^* = 1.105$, $\tilde{\mu}_2^* = -2.084$ and $\tilde{\kappa}^* = 2.989$ with $p(\alpha) = 3\log(1 - |1 - 2\alpha|)$. We obtain $M_n = 18.573$, $IM_n^{(0)} = 16.262$ and $IM_n^{(1)} = 18.542$, all suggesting strong evidence to reject the unicomponent von Mises distribution. In addition, the values of the two estimated mean direction parameters are roughly $\pi$ apart, which suggests that the two subgroups of turtles traveled in opposite directions.
Table 2: Simulated power of the MLRT and the IMLRT with 5% level

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
<th>$IM_n^{(0)}$</th>
<th>$IM_n^{(1)}$</th>
<th>$M_n^{(0)}$</th>
<th>$IM_n^{(0)}$</th>
<th>$IM_n^{(1)}$</th>
<th>$M_n^{(0)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.05</td>
<td>5.8</td>
<td>6.1</td>
<td>6.2</td>
<td>24.0</td>
<td>24.7</td>
<td>24.4</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>11.2</td>
<td>19.7</td>
<td>20.7</td>
<td>70.6</td>
<td>73.0</td>
<td>72.7</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>80.1</td>
<td>84.0</td>
<td>84.1</td>
<td>99.6</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>200</td>
<td>0.05</td>
<td>7.5</td>
<td>8.0</td>
<td>8.1</td>
<td>47.2</td>
<td>54.8</td>
<td>54.8</td>
</tr>
<tr>
<td>200</td>
<td>0.10</td>
<td>37.9</td>
<td>44.8</td>
<td>45.8</td>
<td>97.0</td>
<td>97.5</td>
<td>97.5</td>
</tr>
<tr>
<td>200</td>
<td>0.25</td>
<td>96.4</td>
<td>98.6</td>
<td>99.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

In this paper, we discussed the use of the MLRT and the IMLRT in a two-component von Mises mixture with a structural parameter. Since the concentration parameter tends to be overestimated by the ordinary MLE or the modified MLE under the von Mises mixture model, we proposed a novel penalty on the size of the concentration parameter to avoid the problem of overestimation. We also applied a more effective penalty on the mixing proportion. The limiting distributions of the resulting test statistics were shown to be $\frac{1}{2} \chi^2_0 + \frac{1}{2} \chi^2_1$. Our simulation studies demonstrated that the two methods greatly enhance the accuracy of the tests.

The von Mises distribution is often regarded as normal distribution for circular data. However, the von Mises mixtures have quite different properties from normal mixtures. Most notably, the von Mises mixture is strongly identifiable (Chen 1995). Because of this, the MLRT has quite different asymptotic properties when applied to normal mixtures and von Mises mixture models. For von Mises mixtures, the convergence rates of the MLEs of the mixing distribution and $\kappa$ are $O_p(n^{-1/4})$ and $O_p(n^{-1/2})$ respectively, and the limiting distribution of the MLRT has a simple form, which are different from the conclusions for normal mixtures (Chen & Chen 2003, Chen & Kalbfleisch 2005). At the same time, when $\kappa$ goes to infinity, the von Mises distribution converges to a normal distribution and the strong identifiability of the von Mises distribution weakens, which is reflected in (9) in the proof of Theorem 1. Although for each given $\kappa$, the ratio in (9) remains $o_p(1)$, the asymptotic approximation requires larger $n$ for larger $\kappa$.

Under certain regularity conditions, the results in this paper can be generalized to other parametric kernels, for example, mixtures of Gamma distributions. The generalization is compelling only if it also includes the normal mixture models and more, which is likely technically involved. The research in this direction is underway. Some other questions still need to be addressed in the use of the modified likelihood approach. In particular, we use simulations to determine the level of modification to best compromise between the accuracy of the levels and efficiency of the test. A data-driven procedure with some theoretical justification is warranted.

APPENDIX

Preliminaries and Notation. Define

\[
U_\kappa = \frac{1}{\kappa - \kappa_0} \left\{ \frac{f(\theta_i; 0, \kappa)}{f(\theta_i; 0, \kappa_0)} - 1 \right\} = \frac{1}{\kappa - \kappa_0} \left[ \frac{I_0(\kappa_0)}{I_0(\kappa)} \exp\{(\kappa - \kappa_0) \cos X_i\} - 1 \right],
\]

\[
Y_{\mu, \kappa} = \frac{1}{\mu} \left\{ \frac{f(\theta_i; \mu, \kappa)}{f(\theta_i; 0, \kappa)} - \frac{\mu}{\kappa_0} \right\}
= \frac{I_0(\kappa_0)}{\mu I_0(\kappa)} \exp\{\kappa \cos(\theta_i - \mu) - \kappa_0 \cos \theta_i\} - \exp\{(\kappa - \kappa_0) \cos \theta_i\},
\]

\[
Z_{\mu} = \frac{Y_{\mu}(\mu, \mu_0) - Y_{\mu}(0, \kappa_0)}{\mu},
\]

8
and let $Y_i(0, \kappa), U_i(\kappa_0)$ and $Z_i(0)$ be their continuity limits. For convenience, we put $Y_i(\mu) = Y_i(\mu, \kappa_0), Y_i = Y_i(0), U_i = U_i(\kappa_0)$ and $Z_i = Z_i(0)$. The following proposition assesses the stochastic orders of some relevant stochastic processes.

**Proposition 1.** Indexed by the parameters $\kappa \in [\kappa_0 - \delta, \kappa_0 + \delta]$ for some $\delta > 0$, and $|\mu| \leq \pi$, the following processes are tight

\[
U_n^*(\kappa) = n^{-1/2} \sum (U_i(\kappa) - U_i)/(\kappa - \kappa_0),
\]

\[
Y_n^*(\mu, \kappa) = n^{-1/2} \sum (Y_i(\mu, \kappa) - Y_i(\mu))/(\kappa - \kappa_0),
\]

\[
z_n^*(\mu, \kappa) = n^{-1/2} \sum (Y_i(\mu, \kappa) - Y_i(\mu))/\mu^2.
\]

The tightness of the first three processes was shown in Fu, Chen & Li (2007). The proof of the tightness of the last process is the same.

**Proof of Theorem 1.** Let $r_n(\alpha, \mu_1, \mu_2, \kappa) = r_{1n}(\alpha, \mu_1, \mu_2, \kappa) + r_{2n}$, where $r_{1n}(\alpha, \mu_1, \mu_2, \kappa) = 2\{l_n(\alpha, \mu_1, \mu_2, \kappa) - l_n(0, 0, 0, \kappa_0)\}$ and $r_{2n} = 2\{l_n(0, 0, 0, \kappa_0) - l_n(0, \mu_0, \mu_0, \kappa_0)\}$. Then the LRT statistic $R_n = r_n(\alpha, \mu_1, \mu_2, \hat{\kappa})$.

By symmetry, we assume that $0 \leq \alpha \leq 1/2$ instead of $0 \leq \alpha \leq 1$. In light of Lemma 2, under the null distribution, $\hat{\mu}_1 = o_p(1)$ and $\hat{\mu}_2 = o_p(1)$. Hence, for asymptotic consideration, we only need to expand $r_{1n}$ at $\mu_1$ values in an arbitrarily small neighborhood of 0. Expansion of $r_{1n}$ with respect to $\mu_2$ will be done in

\[
\Omega_1(\epsilon) = \{|\mu_2| > \epsilon\} \text{ and } \Omega_2(\epsilon) = \{|\mu_2| \leq \epsilon\}
\]

for arbitrarily small $\epsilon > 0$, respectively. Let $R_n(\epsilon, I)$ denote the supremum of $r_n$ over $\Omega_1(\epsilon)$ and $R_n(\epsilon, II)$ denote the supremum of $r_n$ over $\Omega_2(\epsilon)$. Then $R_n = \max\{R_n(\epsilon, I), R_n(\epsilon, II)\}$. Since $\hat{\kappa}$ is a consistent estimator of $\kappa_0$ as shown in Lemma 2, we need only expand $r_{1n}$ with respect to $\kappa$ in $[\kappa_0 - \delta, \kappa_0 + \delta]$ for some arbitrarily small $\delta > 0$.

We first analyze $R_n(\epsilon, I)$. Express

\[
r_{1n}(\alpha, \mu_1, \mu_2, \kappa) = 2 \sum_{i=1}^n \log(1 + \delta_i),
\]

where

\[
\delta_i = (\kappa - \kappa_0)U_i(\kappa_0) + (1 - \alpha)\mu_1 Y_i(0, \kappa_0) + \alpha \mu_2 Y_i(\mu_2, \kappa_0) + \epsilon_{in}
\]

\[
= (\kappa - \kappa_0)U_i + (1 - \alpha)\mu_1 Y_i + \alpha \mu_2 Y_i(\mu_2) + \epsilon_{in},
\]

and $\epsilon_{in}$ is the remainder term. Let $\epsilon_n = \sum_{i=1}^n \epsilon_{in}$. By Proposition 1, we can show

\[
|\epsilon_n| \leq n^{1/2} \{(\kappa - \kappa_0)^2 + \alpha^2 + \mu_1^2\} O_p(1).
\]

Since the remainder resulting from the square and cubic sums has at least the order of the remainder
from the linear sum, we have
\[
E_1(n, \alpha, \mu_1, \mu_2, \kappa) \leq 2 \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} \delta_i^2 + \frac{2}{3} \sum_{i=1}^{n} \delta_i^3
\]
\[
= 2 \sum_{i=1}^{n} \{(\kappa - \kappa_0)U_i + (1 - \alpha)\mu_1 Y_i + \alpha \mu_2 Y_i(\mu_2)\}
- \sum_{i=1}^{n} \{(\kappa - \kappa_0)U_i + (1 - \alpha)\mu_1 Y_i + \alpha \mu_2 Y_i(\mu_2))^2
+ n^{1/2}(\kappa - \kappa_0)^2 + \alpha^2 + \mu_1^2)O_p(1) + n((\kappa - \kappa_0)^3 + \alpha^3 + \mu_1^3)O_p(1).
\]
(4)

Let
\[
V_i(\mu_2) = \frac{1}{\mu_2} \left[ Y_i(\mu_2) - \frac{E[Y_i(\mu_2)U_i]}{E(U_i^2)} U_i - \frac{E[Y_i(\mu_2)Y_i]}{E(Y_i^2)} Y_i \right]
= \frac{1}{\mu_2} \left[ Y_i(\mu_2) - \frac{A(\kappa_0)(\cos \mu_2 - 1)}{\mu_2 [1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)]} U_i - \sin \mu_2 Y_i \right]
\]
and \(V_i = V_i(0)\) be the continuity limit of \(V_i(\mu_2)\). Then
\[
(\kappa - \kappa_0)U_i + (1 - \alpha)\mu_1 Y_i + \alpha \mu_2 Y_i(\mu_2) = t_1 U_i + t_2 Y_i + t_3 V_i(\mu_2),
\]
for some \(t_1, t_2\) and \(t_3 = \alpha \mu_2^2\).

It is easy to verify that \(U_i, Y_i\) and \(V_i(\mu_2)\) are mutually orthogonal for all \(\mu_2\). We restrict our attention to a small neighborhood of \((t_1, t_2, t_3) = (0, 0, 0)\) as suggested by the consistency results of the MLEs in Lemma 2. Consequently, we may regard \(t_1, t_2\) and \(t_3\) as \(o_p(1)\). We have
\[
E_1(n, \alpha, \mu_1, \mu_2, \kappa) \leq 2 \sum_{i=1}^{n} \{t_1 U_i + t_2 Y_i + t_3 V_i(\mu_2)\}
- \sum_{i=1}^{n} \{t_1^2 U_i^2 + t_2^2 Y_i^2 + t_3^2 V_i^2(\mu_2)\}\{1 + o_p(1)\}.
\]
(5)

The remainder terms in (4) are summarized in the \(o_p(1)\) in (5). Thus
\[
E_1(n, \alpha, \mu_1, \mu_2, \kappa) \leq \frac{(\sum U_i)^2}{\sum U_i^2} + \frac{(\sum Y_i)^2}{\sum Y_i^2} + \sup_{\epsilon < |\mu_2| \leq \pi} \frac{[(\sum V_i(\mu_2))^2]^2}{\sum V_i^2(\mu_2)} + o_p(1).
\]
(6)

On the other hand, the classic analysis gives
\[
E_2(n, \alpha, \mu_1, \mu_2, \kappa) = \frac{(\sum U_i)^2}{\sum U_i^2} - \frac{(\sum Y_i)^2}{\sum Y_i^2} + o_p(1).
\]
(7)

Combining (6) and (7) yields
\[
R_1(\epsilon, I) \leq \sup_{\epsilon < |\mu_2| \leq \pi} \frac{[(\sum V_i(\mu_2))^2]^2}{\sum V_i^2(\mu_2)} + o_p(1).
\]

We thus obtained a upper bound for \(R_1(\epsilon, I)\). Similar to Chen & Chen (2003), this upper bound is achievable. That is,
\[
R_1(\epsilon, I) = \sup_{\epsilon < |\mu_2| \leq \pi} \frac{[(\sum V_i(\mu_2))^2]^2}{\sum V_i^2(\mu_2)} + o_p(1).
\]
(8)
This concludes the analysis of $R_n(\epsilon, I)$.

Next, we try to expand $R_n(\epsilon, II)$. Due to the arbitrariness of $\epsilon$, in the region $\Omega_2(\epsilon)$, we may regard $\mu_1$, $\mu_2$ and $\kappa - \kappa_0$ as $o_p(1)$. The consequence of this treatment will be spelled out in the due place. We write
\[
\delta_i = (\kappa - \kappa_0)U_i + m_1Y_i + m_2Z_i + \epsilon_{in},
\]
where $\epsilon_{in}$ is the remainder term, $m_1 = (1-\alpha)\mu_1 + \alpha \mu_2$ and $m_2 = (1-\alpha)\mu_1^2 + \alpha \mu_2^2$. Let $\epsilon_n = \sum_{i=1}^{n} \epsilon_{in}$. By Proposition 1, we find
\[
\epsilon_n = n^{1/2}(\kappa - \kappa_0)^2O_p(1) + n^{1/2}m_1(\kappa - \kappa_0)O_p(1)
+ n^{1/2}(1-\alpha)\mu_1^3O_p(1) + n^{1/2}\alpha \mu_2^3O_p(1).
\]
Using the facts $|2x| \leq 1 + x^2$ for any $x$ and $\mu_1 = o_p(1)$, $\mu_2 = o_p(1)$, $\kappa - \kappa_0 = o_p(1)$, we have
\[
|\epsilon_n| \leq n\{n(\kappa - \kappa_0)^2 + m_1^2 + \alpha m_2^2\}o_p(1) + o_p(1).
\]
Note that $n^{-1} \sum_{i=1}^{n} [(\kappa - \kappa_0)U_i + m_1Y_i + m_2Z_i]^2$ converges to a positive definite quadratic form in $\kappa - \kappa_0$, $m_1$ and $m_2$. Thus
\[
\frac{\sum_{i=1}^{n} |(\kappa - \kappa_0)U_i + m_1Y_i + m_2Z_i|^2}{\sum_{i=1}^{n} [(\kappa - \kappa_0)U_i + m_1Y_i + m_2Z_i]^2} \leq O_p(1). \tag{9}
\]
Using a few similar techniques employed for $R_n(\epsilon, I)$, we have
\[
\begin{align*}
\sum_{i=1}^{n} &\{s_1U_i + m_1Y_i + m_2V_i\} \\
&- \sum_{i=1}^{n} \{s_1^2U_i^2 + m_1^2Y_i^2 + m_2^2V_i^2\}(1 + o_p(1)) + o_p(1).
\end{align*}
\]
for some $s_1$ that is linear in $\kappa$ and $m_2$. Therefore,
\[
r_{1n}(\alpha, \mu_1, \mu_2, \kappa) \leq \frac{\{\sum U_i\}^2}{\sum U_i^2} + \frac{\{\sum Y_i\}^2}{\sum Y_i^2} + \frac{\{\sum V_i\}^2}{\sum V_i^2} + o_p(1)
\]
and
\[
r_{2n} = 2\{l_n(0, 0, 0, 0) - l_n(0, \hat{\mu}_0, \hat{\mu}_0, 0)\} = -\frac{\{\sum U_i\}^2}{\sum U_i^2} - \frac{\{\sum Y_i\}^2}{\sum Y_i^2} + o_p(1).
\]
Then
\[
R_n(\epsilon, II) \leq \frac{\{\sum V_i\}^2}{\sum V_i^2} + o_p(1). \tag{10}
\]
If $\mu_1$ and $\mu_2$ are not regarded as $o_p(1)$ quantities but parameters in $\Omega_2(\epsilon)$, the above result will be more precisely stated as
\[
R_n(\epsilon, II) \leq \frac{\{\sum V_i\}^2}{\sum V_i^2} + o_p(1) + \epsilon O_p(1).
\]
See similar arguments in Chen & Chen (2001, 2003). It can also be shown that
\[
R_n(\epsilon, II) \geq \frac{\{\sum V_i\}^2}{\sum V_i^2} + o_p(1). \tag{11}
\]
For any $\epsilon > 0$, $R_n = \max\{R_n(\epsilon, I), R_n(\epsilon, II)\}$. Combining (8) and (11), we have

$$R_n \geq \max \left\{ \frac{[(\sum V_i)^+]^2}{\sum V_i^2}, \sup_{\epsilon < |\mu| \leq \pi} \frac{[(\sum V_i(\mu))^+]^2}{\sum V_i^2(\mu)} \right\} + o_p(1).$$

Note that $\sum V_i(\mu) / \sqrt{\sum V_i^2(\mu)} \xrightarrow{d} \zeta(\mu)$, where $\zeta(\mu)$ is defined in Theorem 1, and $\epsilon > 0$ can be chosen arbitrarily small. Thus, the limit of $R_n$ is stochastically larger than $\sup_{|\mu| \leq \pi} \{\zeta(\mu)\}^2$. It is easily seen that this lower bound is also an asymptotic upper bound. Hence we complete the proof of Theorem 1.

In order to prove Theorem 2, we need the following lemma which states the consistency property of the modified MLEs.

**Lemma 3.** Let $\theta_1, \ldots, \theta_n$ be a random sample from the mixture population $(1 - \alpha)M(\mu_1, \kappa) + \alpha M(\mu_2, \kappa)$. Under the null distribution $M(0, \kappa_0)$,

(a) $\hat{\mu}_0 = o_p(1)$, $\hat{\kappa}_0 - \kappa_0 = o_p(1)$ and

$$pl_n(1/2, \hat{\mu}_0, \hat{\kappa}_0) - pl_n(1/2, 0, 0, \kappa_0) = \frac{[\sum U_i(\epsilon)]^2}{\sum U_i^2} + \frac{[\sum Y_i(\epsilon)]^2}{\sum Y_i^2} + o_p(1);$$

(b) $\hat{\mu}_1 = o_p(1)$, $\hat{\mu}_2 = o_p(1)$ and $\hat{\kappa} - \kappa_0 = o_p(1)$.

The proof of Lemma 3 is given in Chen, Li & Fu (2007).

**Proof of Theorem 2.** (a) By (7), (12) and the consistency of $\hat{\kappa}$, we have

$$M_n = r_n(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}) + r_{2n} + 2[p(\hat{\alpha}) - p(0.5)] + o_p(1)$$

$$= r_n(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}) + 2[p(\hat{\alpha}) - p(0.5)] + o_p(1)$$

$$\leq r_n(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}) + o_p(1).$$

Since $(\hat{\mu}_1, \hat{\mu}_2, \hat{\kappa})$ are consistent estimators of $(0, 0, \kappa_0)$, $R_n(\epsilon, I)$ can serve as an upper bound for $r_n(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\kappa})$. Combining (10) and (13), we have

$$M_n \leq \frac{[\sum V_i^+]^2}{\sum V_i^2} + o_p(1).$$

The above bound is achievable as before. Consequently, the limiting distribution of $M_n$ is given by $\frac{1}{2} \lambda_0^2 + \frac{1}{2} \lambda_1^2$.

(b) Obviously,

$$IM_n^{(K)} \leq M_n \leq \frac{[\sum V_i^+]^2}{\sum V_i^2} + o_p(1).$$

If one of $\alpha_j$’s is equal to 0.5,

$$IM_n^{(K)} \geq \frac{[\sum V_i^+]^2}{\sum V_i^2} + o_p(1).$$

Hence

$$IM_n^{(K)} = \frac{[\sum V_i^+]^2}{\sum V_i^2} + o_p(1).$$

Consequently, the limiting distribution of $IM_n^{(K)}$ is given by $\frac{1}{2} \lambda_0^2 + \frac{1}{2} \lambda_1^2$. 

12
ACKNOWLEDGEMENTS

The research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank the AE and two referees for their valuable suggestions and constructive comments.

REFERENCES


---

Received ???
Accepted ???

Jiahua CHEN: jhchen@stat.ubc.ca
Department of Statistics, University of British Columbia
Vancouver, BC
Canada, V6T 1Z2

Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, ON
Canada, N2L 3G1

Pengfei LI: p4li@uwaterloo.ca
Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, ON
Canada, N2L 3G1

Yuejiao FU: yuejiao@mathstat.yorku.ca
Department of Mathematics and Statistics, York University
Toronto, ON
Canada, M3J 1P3