Sankhyā A : The Indian Journal of Statistics
https://doi.org/10.1007/s13171-023-00337-3

Clustering Circular Data via Finite Mixtures of von Mises Distributions and an Application to Data on Wind Directions

S. Rao Jammalamadaka
University of California, Santa Barbara, USA
V.S. Vaidyanathan
Pondicherry University, Pondicherry, India

Abstract

The von Mises distribution, which is also known as the Circular Normal distribution is a well-studied and commonly used distribution for analyzing data on a unit circle. It has many properties and similarities to the normal distribution defined on the real line, making it popular for modeling circular data. Since it is unimodal, finite mixtures of von Mises distributions may be used to deal with circular data that may potentially have more than one mode. In this paper, our goal is to cluster such data sets after approximating each data set as a finite mixture of von Mises distributions. To accomplish such clustering we need a distance measure between any two such finite mixtures. For this, we propose using the Kullback-Liebler and Bhattacharyya distance measures. The applicability and usefulness of the proposed measures in identifying clusters present in a data set is first demonstrated through a simulation study. A real-life application that clusters the surface wind direction data in five major Indian cities is then studied using the proposed measures.

AMS (2000) subject classification. 62R10, 62H11.
Keywords and phrases. Bhattacharyya distance, hierarchical clustering, Kullback-Liebler divergence, von Mises mixture distribution, matching based bound, surface wind direction..

1 Introduction

Directional data represent observations that are directions in two, three, or higher dimensions, recorded on a unit circle, the unit sphere, or high-dimensional manifolds respectively. Examples of such data include wind directions, direction of earth's magnetic field, etc. Analyses of such data should employ metrics that take into account the corresponding topological space, instead of the metrics one uses on Euclidean spaces. This novel area

Published online: 03 January 2024

has been studied in texts like Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), Fisher (1995) etc., and is the subject of active current research. Among probability distributions defined on a unit circle, the von Mises (vM) distribution, also called the Circular Normal (CN) distribution, is the most commonly used distribution to model circular data. This is a two parameter distribution with the probability density function (pdf) (see e.g. Jammalamadaka and SenGupta 2001)

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta-\mu)}, \quad 0 \le \theta < 2\pi,$$

where $0 \le \mu < 2\pi$ and $\kappa \ge 0$ denote respectively the mean direction and concentration parameters, while $I_0(\kappa)$ is the modified Bessel function of the first kind and of order zero. The vM distribution is symmetric about the mean direction μ, and is unimodal with the mode at $\theta = \mu$. Large values of κ indicate higher concentration around μ. As stated in Jammalamadaka et al. (2021), any probability distribution defined on unit circle can be expressed as a countable mixture of vM distributions, while it can be *approximated* by a finite mixture with sufficiently large number of components. Thus, a finite mixture of vM distributions can be used to model circular data having more than one mode. The pdf of a mixture of k-von Mises distributions labelled $vMmix(k)$, with mixing proportions p_i and parameters $(\mu_i, \kappa_i), i = 1, 2, \ldots, k$ is given by

$$g(\theta) = \sum_{i=1}^{k} p_i \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(\theta-\mu_i)}, \quad 0 \le \theta < 2\pi, \tag{1}$$

where $p_i > 0, i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. Estimation of the parameters and the mixing proportions can be done using the expectation-maximization (EM) algorithm—see e.g. Dhillon and Sra (2003) and Banerjee et al. (2009). Hornik and Grün (2014) implements the EM algorithm in the R statistical package movMF. As an example, we consider the turtle data set mentioned in Stephens (1969), which shows the orientation of 76 turtles after laying eggs. This is shown in Fig. 1 as a circular plot.

It is seen from this plot that there is more than one preferred direction and hence the data can be modelled using Eq. 1. A fit of Eq. 1 with $k = 2$ for the data using movMF() function results in $\mu_1 = 241.2036°; \mu_2 = 63.4716°; \kappa_1 = 8.4465; \kappa_2 = 2.6187; p_1 = 0.16; p_2 = 0.84$. Calderara et al. (2011) has developed a novel approach for classifying people trajectories by describing the trajectories as sequence of angles and modelling them through

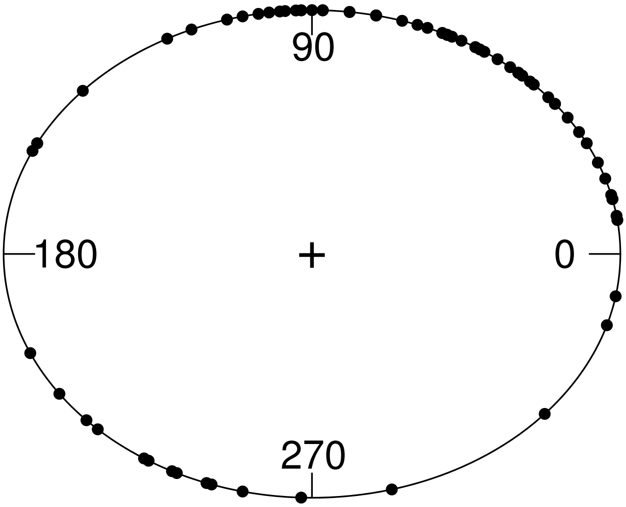

Figure 1: Circular plot of orientation of turtles

finite mixture of vM distributions. Using finite mixtures of vM distributions, Masseran et al. (2013) models the average hourly wind direction data for nine wind stations located in Peninsular Malaysia. Jammalamadaka et al. (2021) compares the performance of Euclidean and Kullback-Liebler (KL) distance between finite mixtures of vM distributions. As they state, although there is an explicit expression for the KL distance between any single vM model from any other vM model, there is no such expression for the KL distance between two finite mixtures of vM distributions, and they then use numerical integration techniques. In this paper, we propose a matching based lower bound for the KL divergence between two $vMmix(k)$ distributions. Also, a close upper bound for the Bhattacharyya divergence between two $vMmix(k)$ distributions is proposed. These divergence measures can then be utilized to cluster different sets of circular data each of which is approximated by $vMmix(k)$ distributions. The rest of the paper is organized as follows. In Section 2, the motivation for the research problem is explained and the KL and the Bhattacharyya divergence measures are defined. Bounds to KL and Bhattacharyya divergence (labelled the "B divergence" from now on) measures to compute the distance between two $vMmix(k)$ distributions are introduced in Section 3. Numerical illustration highlighting the application of KL and B divergence measures in clustering observations from different $vMmix(k)$ models is presented in Section 4 through an extensive simulation study. Application of the proposed $vMmix(k)$ distribution to model the surface wind direction data of five Indian cities and identifying clusters present in them, based both on the KL and B divergence, is the topic of Section 5. We end with brief concluding remarks in Section 6.

2 Motivation and Methodology

Wind energy is a renewable energy that is gaining popularity as a method used to generate electricity. This is done through wind turbines that convert the energy from the wind. Positioning the wind turbines in alignment with the wind direction is very crucial for the optimal generation of electricity. A proper planning of where to install wind turbines and how many to install in a given geographical region is necessary for the optimal generation of wind energy. This involves identifying geographical regions that have similar wind directions and adopting a plan of action for each group. However, the wind directions in a geographical region is usually not oriented to a single direction and it changes during different time points $t_1, t_2, \ldots, t_p$ of the day. Thus, the wind directions can be effectively modeled using a finite mixture of circular distributions defined on a unit circle. Further, grouping of the geographical regions based on the wind directions can be made by computing the distances between the mixture distributions. Considering the circular distribution to be vM, the problem is to cluster, $vMmix(k)$ distributions in some L locations, where the number of mixture components need not be the same for each location. The data template is shown in Table 1.

The main focus of this paper is to define the divergence between two $vMmix(k)$ mixture-distributions and use them to obtain clusters in circular data that might be multimodal. Towards this goal, bounds to the KL and B divergence measures between any two vM mixtures, are proposed.

2.1 KL and B Divergence Measures Let f and g denote two probability density functions defined on the real line.

1. Kullback-Leibler (KL) divergence (Kullback and Leibler 1951) between two densities f and g is defined as

$$\begin{aligned} KL(f,g) &= \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \\ &= E_f\left(\log\left(\frac{f(x)}{g(x)}\right)\right). \end{aligned}$$

   It is also known as relative entropy and measures the degree of separation of f and g. KL is non-negative and becomes zero only when

Table 1: Data template

Location	Wind directions	Mixture density
1	$\theta_{t_1}, \theta_{t_2}, \ldots, \theta_{t_p}$	$vM(\mu_i, \kappa_i), i = 1, 2, \ldots, k_1.$
2	$\theta_{t_1}, \theta_{t_2}, \ldots, \theta_{t_p}$	$vM(\mu_i, \kappa_i), i = 1, 2, \ldots, k_2.$
$\vdots$	$\vdots$	$\vdots$
L	$\theta_{t_1}, \theta_{t_2}, \ldots, \theta_{t_p}$	$vM(\mu_i, \kappa_i), i = 1, 2, \ldots, k_L.$

$f = g$. However, it is not a symmetric measure. Jammalamadaka et al. (2021) has derived the expression for the KL divergence between two vM distributions with respective densities, $f(\mu_1, \kappa_1)$ and $g(\mu_2, \kappa_2)$, and it is given by

$$KL(f, g) = \log(I_0(\kappa_2)) - \log(I_0(\kappa_1)) + \kappa_1 A(\kappa_1) - \kappa_2 \cos(\mu_1 - \mu_2) A(\kappa_1), \qquad (2)$$

where $A(\kappa) = \dfrac{I_1(\kappa)}{I_0(\kappa)}, I_1(\kappa) = \dfrac{\partial}{\partial \kappa} I_0(\kappa)$.

2. Bhattacharyya Coefficient (Bhattacharyya 1943) between f and g is defined as

$$\begin{aligned} CB(f, g) &= \int_{-\infty}^{\infty} \sqrt{f(x) g(x)} dx \qquad (3) \\ &= E_f \left(\sqrt{\frac{g(x)}{f(x)}} \right). \end{aligned}$$

$CB(f, g)$ measures the degree of overlap between f and g. It has the following properties.

- $0 \leq CB \leq 1$.
- CB=1 when $f = g$.
- CB is symmetric.

For two vM distributions with respective densities, $f(\mu_1, \kappa_1)$ and

$g(\mu_2, \kappa_2)$, the expression for $CB(f, g)$ is given by Calderara et al. (2011)

$$CB(f,g) = \left[\sqrt{\frac{1}{I_0(\kappa_1) I_0(\kappa_2)}} I_0 \left(\frac{\sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(\mu_1 - \mu_2)}}{2} \right) \right]. \quad (4)$$

The Bhattacharyya divergence between f and g which we will denote by $B(f, g)$ is defined as

$$B(f, g) = -\log(CB(f, g)),$$

where $CB(f, g)$ is as given in Eq. 4.

Figure 2 depicts the line plots of the divergence between $f(\mu_1, \kappa_1)$ and $g(\mu_2, \kappa_2)$. The divergence measures are computed taking $\kappa_1 = 1, \kappa_2 = 2$ and varying the difference between the mean directions μ_1 and μ_2.

One should keep in mind that this expression for $B(f, g)$ as well as the expression for $KL(f, g)$ given in Eq. 2 are for single vM components. We now deal with such divergence measures for mixtures.

3 KL Divergence Between 2 vM Mixtures

Let $\mathbf{f} = \sum_{i=1}^{p} \alpha_i f_i$ and $\mathbf{g} = \sum_{j=1}^{q} \beta_j g_j$ be respectively a $vMmix(p)$ and a $vMmix(q)$ distribution, where $f_i = vM(\mu_i, \kappa_i), i = 1, 2, \ldots, p$ and

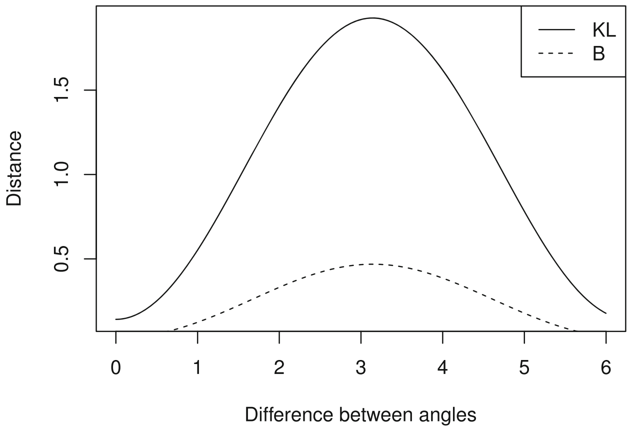

Figure 2: Line plots of KL and B divergence measures between two vM distributions

$g_j = vM(\mu_j, \kappa_j), j = 1, 2, \ldots, q$. Here the $\{\alpha_i\}$ and $\{\beta_j\}$ denote the mixing proportions such that $\sum_{i=1}^{p} \alpha_i = 1$ and $\sum_{j=1}^{q} \beta_j = 1$. The KL divergence between such $\mathbf{f}$ and $\mathbf{g}$ is given by

$$\begin{aligned} KL(\mathbf{f}, \mathbf{g}) &= E_{\mathbf{f}} \left(\log \left(\frac{\mathbf{f}}{\mathbf{g}} \right) \right) \\ &= E_{\mathbf{f}} \left(\log \left(\frac{\sum_{i=1}^{p} \alpha_i f_i}{\sum_{j=1}^{q} \beta_j g_j} \right) \right). \end{aligned}$$

Similarly, the Bhattacharyya coefficient between $\mathbf{f}$ and $\mathbf{g}$ is given by

$$CB(\mathbf{f}, \mathbf{g}) = E_{\mathbf{f}} \left(\sqrt{\frac{\sum_{i=1}^{p} \alpha_i f_i}{\sum_{j=1}^{q} \beta_j g_j}} \right).$$

Since the vM distribution is not closed under addition, an explicit expression for $KL(\mathbf{f}, \mathbf{g})$ and $CB(\mathbf{f}, \mathbf{g})$ is not available for these mixtures. However, they can be obtained using fairly accurate approximations. We do this by getting a matching based bound for $KL(\mathbf{f}, \mathbf{g})$ and an upper bound for $CB(\mathbf{f}, \mathbf{g})$.

3.1 Matching Based Bound for $KL(\mathbf{f}, \mathbf{g})$ This type of bound has been originally proposed and studied in Goldberger and Gordon (2003) to compute the KL divergence between two Gaussian mixtures, in a very similar context. This bound is based on the idea that one particular component in $\mathbf{g}$ is closer (dominates) than all the other components of $\mathbf{f}$. In the context of divergence between mixtures of two circular distributions, the matching based bound for KL divergence assumes that an arc (say component 1) in the first mixture distribution is closer to the other arcs of the second mixture distribution. As an illustration, consider the rose diagram of observations from two mixture densities $\mathbf{f}$ and $\mathbf{g}$ depicted in Fig. 3. It can be observed from Fig. 3 that one petal in $\mathbf{g}$ is closer to (overlaps) all the other petals in $\mathbf{f}$. Thus, the matching based bound for $KL(\mathbf{f}, \mathbf{g})$ is computed based on the component in $\mathbf{g}$ that is closer to the components in $\mathbf{f}$.

Thus

$$\begin{aligned} KL(\mathbf{f}, \mathbf{g}) &= \int \mathbf{f} \log \left(\frac{\mathbf{f}}{\mathbf{g}} \right) d\theta \\ &= \int \sum_{i=1}^{p} \alpha_i f_i \log \left(\frac{\mathbf{f}}{\mathbf{g}} \right) d\theta \end{aligned}$$

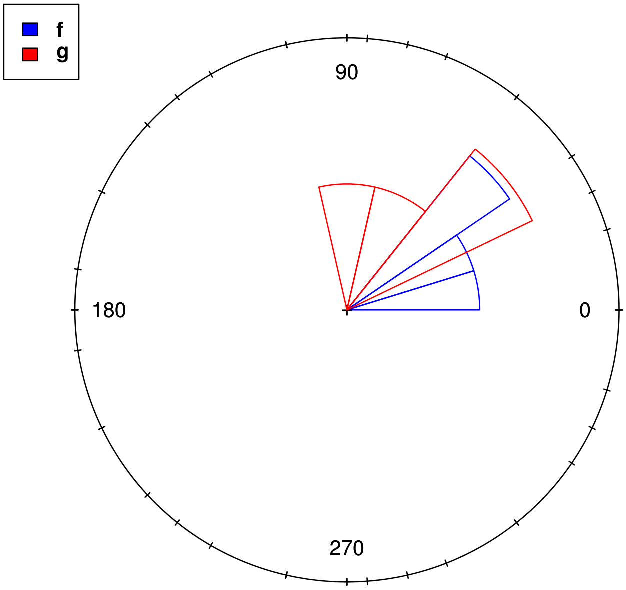

Figure 3: Superimposed rose plot from two mixture densities

$$
\begin{aligned}
&= \sum_{i=1}^{p} \alpha_i \int f_i \log(\mathbf{f}) d\theta - \sum_{i=1}^{p} \alpha_i \int f_i \log(\mathbf{g}) d\theta \\
&= \sum_{i=1}^{p} \alpha_i \int f_i \log\left(\sum_{i=1}^{p} \alpha_i f_i\right) d\theta - \sum_{i=1}^{p} \alpha_i \int f_i \log\left(\sum_{j=1}^{q} \beta_j g_j\right) d\theta \\
&\geq \sum_{i=1}^{p} \alpha_i \int f_i \log(\alpha_i f_i) d\theta - \sum_{i=1}^{p} \alpha_i \max_j \int f_i \log(\beta_j g_j) d\theta \\
&= \sum_{i=1}^{p} \alpha_i \int f_i \log(\alpha_i f_i) d\theta + \sum_{i=1}^{p} \alpha_i \min_j \int -[f_i \log(\beta_j g_j) d\theta] \\
&= \sum_{i=1}^{p} \alpha_i \min_j \left(KL(f_i, g_j) + \log\left(\frac{\alpha_i}{\beta_j}\right)\right).
\end{aligned}
$$

Define $\pi(i) = \arg\min_j \left(KL(f_i, g_j) - \log(\beta_j)\right)$, where $\pi(i)$ denotes the matching function. The matching based bound for $KL(\mathbf{f}, \mathbf{g})$ is given by

$$
KL(\mathbf{f}, \mathbf{g}) \geq \sum_{i=1}^{p} \alpha_i \left(KL(f_i, g_{\pi(i)}) + \log\left(\frac{\alpha_i}{\beta_{\pi(i)}}\right)\right).
$$

We define the matching based KL divergence (KL_{match}) between $\mathbf{f}$ and $\mathbf{g}$ as

$$KL_{match}(\mathbf{f}, \mathbf{g}) = \sum_{i=1}^{p} \alpha_i \left(KL(f_i, g_{\pi(i)}) + \log\left(\frac{\alpha_i}{\beta_{\pi(i)}}\right)\right). \tag{5}$$

We then use the expression for the KL divergence between any two vM distributions given in Eq. 2 to compute $KL(f_i, g_{\pi(i)})$ in Eq. 5. This provides an efficient and easy way to compute such KL divergence instead of resorting to completely numerical methods. However, deriving a mathematical expression for the approximating error bound of (KL_{match}) is difficult. But the accuracy of (KL_{match}) to the true KL divergence (KL_{true}) can be examined by considering the difference between them. Following (Goldberger and Gordon 2003), we define the error between (KL_{true}) and (KL_{match}) as

$$error = \frac{|KL_{true} - KL_{match}|}{KL_{true}}.$$

Since vM distribution is not closed under addition, the true KL divergence between $\mathbf{f}$ and $\mathbf{g}$ is approximated through Monte-Carlo integration as

$$\begin{aligned} KL_{true}(\mathbf{f}, \mathbf{g}) &= \int \mathbf{f} \log\left(\frac{\mathbf{f}}{\mathbf{g}}\right) d\theta \\ &\approx \frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{\mathbf{f}(\theta_\mathbf{i})}{\mathbf{g}(\theta_\mathbf{i})}\right), \end{aligned}$$

where $\theta_1, \theta_2, \ldots, \theta_N$ are sampled from $\mathbf{f}$ and N denote the number of samples. KL_{match} is computed using Eq. 5. The error is computed for the various mixture Cases given in Table 2 and are reported in Table 3.

From Table 3, it is seen that the error between the KL_{true} and KL_{match} is small. Thus, the proposed matching based KL divergence yield values that are closer to the true KL divergence measure.

Suppose the component distributions in $\mathbf{g}$ are identical except for their mixing proportions. In this case, the matching function $\pi(i) = \arg\min_j (KL(f_i, g_j) - \log(\beta_j))$ is the minimum for the component that has the highest mixing proportion β_j. This is because the contribution of $\log(\beta_j)$ to $\pi(i)$ becomes smaller as β_j approaches one. In this sense, the choice of a matching function is typically unique.

Table 2: von Mises mixtures

Mixture	$vMmix(2)$	$vMmix(3)$
Case		
1	$0.50vM(3.14, 3) + 0.50vM(5.25, 3)$	$0.33vM(0, 5) + 0.33vM(2.09, 5) + 0.34vM(4.19, 5)$
2	$0.50vM(3.14, 3) + 0.50vM(5.25, 3)$	$0.33vM(0, 4) + 0.33vM(2.09, 4) + 0.34vM(4.19, 4)$
3	$0.50vM(3.14, 3) + 0.50vM(4.19, 3)$	$0.33vM(0, 5) + 0.33vM(2.09, 5) + 0.34vM(4.19, 5)$
4	$0.25vM(3.14, 3) + 0.75vM(5.25, 3)$	$0.20vM(0, 5) + 0.60vM(2.09, 5) + 0.20vM(4.19, 5)$

Table 3: Error between true KL and matching based KL

Mixture		KL_{true}	KL_{match}	Error
Case	1	2.4872	2.5054	0.0073
	2	2.0373	2.0345	0.0013
	3	3.9025	1.5024	0.6150
	4	2.7455	2.8625	0.0426

3.2 Bhattacharyya Divergence Between 2 vM Mixtures Consider now the B divergence between the finite mixtures $\mathbf{f}$ and $\mathbf{g}$, defined by $B(\mathbf{f},\mathbf{g}) = -\log(CB(\mathbf{f},\mathbf{g}))$. Since

$$\begin{aligned} CB(\mathbf{f},\mathbf{g}) &= \int \sqrt{\mathbf{fg}}\ d\theta \\ &= \int \sqrt{\Big(\sum_{i=1}^{p} \alpha_i f_i\Big)\Big(\sum_{j=1}^{q} \beta_j g_j\Big)}\ d\theta \\ &= \int \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{q} \alpha_i \beta_j f_i g_j}\ d\theta \\ &\le \int \sum_{i=1}^{p}\sum_{j=1}^{q} \sqrt{\alpha_i\beta_j}\sqrt{f_i g_j}\ d\theta \qquad (\because \sqrt{a_1+\ldots+a_n} \le \sqrt{a_1}+\ldots+\sqrt{a_n}) \\ &\le \int \sum_{i=1}^{p}\sum_{j=1}^{q} \sqrt{f_i g_j}\ d\theta \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q} \int \sqrt{f_i g_j}\ d\theta \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q} CB(f_i, g_j). \end{aligned}$$

Thus, $CB(\mathbf{f},\mathbf{g}) \le \sum_{i=1}^{p}\sum_{j=1}^{q} CB(f_i, g_j)$. If and when this upper bound exceeds the value one, it can be scaled back below one by dividing with pq. Therefore,

$$\begin{aligned} B(\mathbf{f},\mathbf{g}) &= -\log(CB(\mathbf{f},\mathbf{g})) \\ &\le -\log\Big(\sum_{i=1}^{p}\sum_{j=1}^{q} CB(f_i, g_j)\Big), \end{aligned} \tag{6}$$

so that Eq. 6 gives an upper bound for the B divergence between two vM mixture distributions. $CB(f_i, g_j)$ in Eq. 6 is computed using Eq. 4.

4 A Simulation Study

In this section, a simulation study is carried out to illustrate the computation of the proposed KL and B divergence between finite mixtures of vM distributions, and compare their effectiveness in identifying clusters. Observations from mixture of vM distributions are simulated for varying number of mixture components and different parameter choices as discussed in Jammalamadaka et al. (2021) under four cases. In each case, two finite mixtures of vM distributions namely, $vMmix(2)$ and $vMmix(3)$ are considered. The parameter choice considered for the mixture distributions under each case is given in Table 2.

For each mixture distribution, samples of size 100 are drawn from each of the component vM distributions. The parameters of the mixture distributions are estimated using the maximum likelihood (ML) method. This process is repeated for $n = 5$ times. Thus, under each Case, there are 10 estimated mixture distributions. These 10 estimated mixture distributions are labelled 1 and 2 respectively based on whether they are $vMmix(2)$ or $vMmix(3)$ distributions. Let $\mathbf{m} = (m_1, m_2, \ldots, m_{10})$ denote these labels arranged such that the first five elements correspond to $vMmix(2)$ and the last five correspond to $vMmix(3)$. i.e., $\mathbf{m} = (1, 1, 1, 1, 1, 2, 2, 2, 2, 2)$. movMF() function in R software is used to simulate the observations from vM mixture distribution and to estimate the parameters. Methodology as given in Banerjee et al. (2009) is used to obtain the ML estimates of the parameters. For each Case, the KL and B divergence between the estimated mixture distributions are computed using KL_{match} and the Bhattacharyya upper bound (scaled to one) given respectively in Eqs. 5 and 6. Based on the divergence values, hierarchical clustering of the 10 mixture distributions under each Case is made. The distances between clusters are computed using the complete linkage method. A reasonable way to identify the number of clusters from hierarchical clustering is to cut the dendrogram at various heights and determine the total within sum of squares of the identified clusters. The final number of clusters is the one for which the total within sum of squares is the smallest. Fixing the required number of clusters as two from the dendrogram of hierarchical clustering, the cluster membership of the mixture distributions based on the KL and B divergence is displayed for each case in Table 4.

Table 4: Cluster membership based on hierarchical clustering

Case		1		2		3		4	
Mixture distribution	Label	KL	B	KL	B	KL	B	KL	B
$vMmix(2)$	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1
$vMmix(3)$	2	2	1	2	2	2	2	2	2
	2	2	2	1	2	2	2	2	2
	2	2	1	2	2	2	2	2	2
	2	1	1	2	1	2	2	2	2
	2	2	1	1	2	2	2	2	2

Matching the actual labels of the mixture distribution with the labels associated with the cluster membership given in Table 4, it is observed that the clusters obtained using the KL and B divergence measures are similar for $vMmix(2)$ mixture distributions. However, there are mismatches in the cluster membership of $vMmix(3)$ mixture distributions. The total number of mismatches in the cluster membership obtained through KL and B divergence are found to be 3 and 5 respectively. Thus, the proposed KL divergence using matching based bound performs considerably better than B divergence in identifying the clusters.

Suppose the parameters of the component vM distributions in $\mathbf{f}$ are close. To see the effectiveness of the proposed KL divergence in identifying the true clusters in such scenario, the simulation study is repeated for the Cases given in Table 5 using the proposed matching based KL divergence.

The corresponding dendrogram is displayed in Fig. 4. It is observed that as the components in $vMmix(2)$ become closer, some observations are misclassified as can be seen from the dendrogram for Cases a, b, and c.

5 A Practical Illustration On Wind Directions

In this section, an implementation of the proposed KL and B divergence measures to compute the distance between mixtures of vM distributions and to detect clusters is illustrated through a real-life data. The data relate to hourly surface wind direction available month wise for each day for various districts

Table 5: $vMmix(2)$ component distributions with closer means

Mixture		$vMmix(2)$	$vMmix(3)$
Case	a	$0.50vM(3.14,3)+0.50vM(3.19,3)$	$0.33vM(0,5)+0.33vM(2.09,5)+0.34vM(4.19,5)$
	b	$0.50vM(3.14,3)+0.50vM(3.24,3)$	$0.33vM(0,5)+0.33vM(2.09,5)+0.34vM(4.19,5)$
	c	$0.50vM(3.14,3)+0.50vM(3.29,3)$	$0.33vM(0,5)+0.33vM(2.09,5)+0.34vM(4.19,5)$
	d	$0.50vM(3.14,3)+0.50vM(3.34,3)$	$0.33vM(0,5)+0.33vM(2.09,5)+0.34vM(4.19,5)$

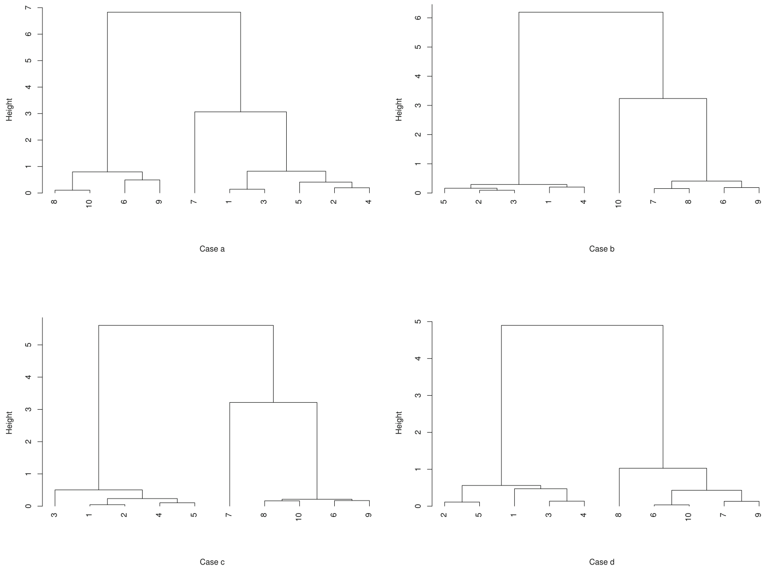

Figure 4: Dendrogram based on matching based KL divergence where $vMmix(2)$ component distributions are close

and cities of India during the year 2015. The data set is available in https://urbanemissions.info/blog-pieces/india-meterology-bydistrict/. For the illustration, the surface wind direction related to five cities namely, Chennai, Visakhapatnam, Trivandrum, Mumbai and Kolkata is considered. The cities Chennai, Visakhapatnam and Kolkata are on the coastal line of the Bay of Bengal, and the cities Trivandrum and Mumbai are on the coastal line of Arabian sea. From the data, the following observations are made.

- In Chennai, between 00:00 to 05:00 hours, the surface wind direction is concentrated in the interval $(0^0, 90^0)$ and $(180^0, 300^0)$, between 06:00 hours to 08:00 hours it is spread over all the directions. After 08:00 hours till mid-night, the direction of wind drifts towards $(0^0, 250^0)$.

- In Visakhapatnam, between 00:00 to 05:00 hours, and between 19:00 hours to mid-night, the wind direction is in the interval $(0^0, 90^0)$, and $(200^0, 360^0)$. During the remaining hours of the day, the wind direction is concentrated in the interval $(90^0, 200^0)$.

- In Trivandrum, throughout the day, the wind direction is more concentrated between $(0^0, 90^0)$, and $(200^0, 300^0)$.
- In Mumbai, between 00:00 to 06:00 hours and between 20:00 hours to mid-night, the predominant wind direction is in the interval $(0^0, 250^0)$. Between 06:00 to 20:00 hours, it is spread in the interval $(250^0, 360^0)$.
- In Kolkata, the wind direction is spread in the interval $(0^0, 360^0)$ between 00:00 to 10:00 hours, and between 20:00 hours to mid-night. Between 10:00 to 20:00 hours, it is relatively more concentrated in the interval $(100^0, 250^0)$.

Thus the surface wind directions in the cities vary considerably over the hours of the day. The rose plot for the wind directions for the above regions is shown in Fig. 5. It can be observed from the rose diagram that the predominant surface wind directions for Chennai are Northeast and South, for Visakhapatnam, Northeast, Southeast and West, for Trivandrum, Northeast and West, for Mumbai, Northwest and Southwest, and for Kolkata, North and South.

Since the surface wind directions for the cities have different orientations, it would be of interest to cluster the cities based on the surface wind directions. Because the cities have more than one predominant surface wind direction, the surface wind direction data for each city is modelled through mixture of von Mises distributions. The estimates of the mixing proportions and the parameters of component mixture densities are obtained through likelihood estimation using the movMF() function. The number of mixture components in the model is varied from 2 to 4. The best fitted mixture model for each city is identified based on the Akaike Information Criterion (AIC) value. Table 6 presents the best fitted mixture model for each city and the ML estimates of the mixing proportions and the parameters of the component von Mises distributions.

From Table 6, it is seen that the surface wind directions of Chennai and Kolkata have a $vMmix(3)$ mixture model whereas that of Visakhapatnam, Trivandrum and Mumbai each have a $vMmix(4)$ mixture model. To cluster the cities based on the estimated mixture models, KL and B divergence (scaled to one) are computed using Eqs. 5 and 6 respectively. Hierarchical clustering with complete linkage method is performed using the computed KL and B divergence. The dendrogram of the respective hierarchical clustering is shown in Fig. 6.

To judge whether the clusters obtained through the KL and B divergence measures are similar, the cophenetic correlation between them is computed using the R package **clue** (Hornik 2005) and is found to be 0.1715. Since the correlation is low, it can be concluded that the clusters obtained through the KL and B divergence measures are not similar. However, comparing the

Figure 5: Rose plot of wind direction data

Table 6: Fit of von Mises mixture distributions

City	Mixture	μ	κ	Mixing proportion
Chennai	$vMmix(3)$	-2.6893 (206°)	0.3197	0.28
		0.7258 (42°)	11.5749	0.32
		-3.1338 (180°)	1.8746	0.4
Visakhapatnam	$vMmix(4)$	-2.6220 (210°)	3.0649	0.25
		1.9492 (112°)	1.6138	0.2
		0.8658 (50°)	5.2434	0.35
		-1.4849 (275°)	4.2588	0.2
Trivandrum	$vMmix(4)$	-1.3932 (280°)	10.8419	0.33
		0.9883 (57°)	71.9861	0.19
		0.8126 (47°)	0.9446	0.21
		-2.1305 (238°)	3.4069	0.27
Mumbai	$vMmix(4)$	-2.0974 (240°)	16.6858	0.18
		-0.6694 (322°)	4.2986	0.27
		-2.2368 (232°)	1.3459	0.22
		0.8114 (46°)	2.3341	0.33
Kolkata	$vMmix(3)$	-2.8410 (197°)	20.1266	0.18
		-0.0584 (357°)	3.0920	0.37
		-2.9206 (193°)	0.9471	0.45

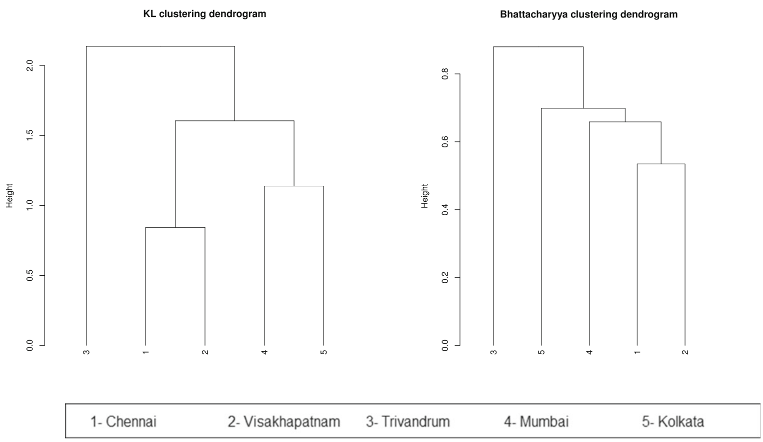

Figure 6: Dendrogram based on the proposed KL and Bhattacharyya measures

rose plot in Fig. 5 and the dendrogram in Fig. 6, it can be observed that the clusters obtained using the matching based KL divergence is fairly good when compared to that of B divergence. Thus, the five cities can be grouped into three clusters, namely, Chennai and Visakhapatnam (cluster 1), Kolkata and Mumbai (cluster 2), and Trivandrum (cluster 3).

6 Concluding Remarks

A methodology that approximates the KL and Bhattacharyya divergence measures to compute the distance between finite mixtures of von Mises distributions is proposed. The matching based method yields a lower bound for the KL divergence between mixtures of vM distributions, while a close upper bound for the Bhattacharyya divergence between such mixtures is provided. The performance of the KL and Bhattacharyya divergence measures is then compared through a simulation study in the context of hierarchical clustering. From this careful simulation study, it is found that the matching based KL divergence performs better in detecting clusters when compared to the Bhattacharyya divergence. A real-life application to clustering the surface wind directions in 5 Indian cities using the proposed KL and Bhattacharyya divergence is highlighted.

The main novelty of the paper is that it addresses the applicability of the KL and Bhattacharyya divergence measures between finite mixtures of von Mises distributions, to provide a basis for clustering of circular data that may be multimodal. Computing these divergence measures is challenging because the von Mises distribution is not closed under addition and thus an explicit expression for the divergence between any finite mixtures of von Mises distributions, is not available. In this paper, bounds for the divergence measures are derived and used to make such computations. The alternate method of employing numerical integration to evaluate the divergences between von Mises mixtures, as done in Jammalamadaka et al. (2021), requires considerably more computational effort as well as large data sets for better accuracy, unlike the methods proposed in this work.

Acknowledgements. The authors thank the Associate Editor and a referee whose knowledgeable comments helped clarify several points and improve the presentation. The research of the second author is supported by DST-SERB, Government of India under MATRICS scheme (File number MTR/2021/000344).

Author contributions Both the authors contributed equally to the work.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

References

Mardia, K.V. and Jupp, P.E. (2000). Directional Statistics. John Wiley & Sons, New Jersey, USA.

Jammalamadaka, S.R. and SenGupta, A. (2001). Topics in Circular Statistics, vol. 5. World Scientific, Singapore.

Fisher, N.I. (1995). Statistical Analysis of Circular Data. Cambridge University Press, Cambridge, UK.

Jammalamadaka, S.R., Wainwright, B. and Jin, Q. (2021). Functional clustering on a circle using von Mises mixtures. *J. Stat. Theory Pract.*, **15**, 1–17.

Dhillon, I.S. and Sra, S. (2003). Modeling data using directional distributions. Technical report, Technical report TR-03-06, Department of Computer Sciences, University of Texas at Austin.

Banerjee, A., Dhillon, I., Ghosh, J. and Sra, S. (2009). Text clustering with mixture of von Mises-Fisher distributions, pp. 151–184. Chapman and Hall/CRC Data mining and Knowledge discovery series, CRC Press.

Hornik, K. and Grün, B. (2014). movmf: An R package for fitting mixtures of von Mises-Fisher distributions. *J. Stat. Softw.*, **58**(10), 1–31.

Stephens, M.A. (1969). Techniques for directional data. Technical report, Department of Statistics, Stanford University, CA.

Calderara, S., Prati, A., Cucchiara, R. (2011). Mixtures of von Mises distributions for people trajectory shape analysis. *IEEE Trans. Circ. Syst. Video Technol.*, **21**(4), 457–471.

Masseran, N., Razali, A.M., Ibrahim, K. and Latif, M.T. (2013). Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia. *Energy Conversion Manag.*, **72**, 94–102.

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Annal Math. Stat.*, **22**(1), 79–86.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin Calcutta Math. Soc.*, **35**, 99–110.

Goldberger, J. and Gordon, S. (2003). An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. In: ICCV, vol. 3, pp. 487–493

Hornik, K. (2005). A CLUE for cluster ensembles. *J. Stat. Softw.*, **14**, 1–25.

S. Rao Jammalamadaka
Department of Statistics and Applied Probability, University of California, Santa Barbara
California, USA
E-mail: sreenivas@ucsb.edu

V.S. Vaidyanathan
Department of Statistics,
Pondicherry University,
Pondicherry Puducherry, India
E-mail: vaidya.stats@pondiuni.ac.in

Paper received: 6 May 2023