Journal of Statistical Theory and Practice (2021) 15:38
https://doi.org/10.1007/s42519-021-00173-4

ORIGINAL ARTICLE

Functional Clustering on a Circle Using von Mises Mixtures

S. Rao Jammalamadaka[1] · Brian Wainwright[1] · Qianyu Jin[2]

Accepted: 25 January 2021

Abstract

This paper addresses the question of clustering density curves around a unit circle by approximating each such curve by a mixture of an appropriate number of von Mises distributions. This is done first by defining a distance between any two such curves either via L^2 or a symmetrized Kullback–Leibler divergence. We show that both these measures yield similar results. After demonstrating via simulations that the proposed clustering methods work successfully, they are applied on an illustrative sample of Optical Coherence Tomography data.

Keywords Circular curves · Von mises (vM) distribution · Mixtures of vM · L^2 Distance and symmetrized Kullback–Leibler · Clustering curves · Optical coherence tomography

Mathematics Subject Classification 62R10 · 62H11

Communicated by Ravi Khattree, Sreenivasa Rao Jammalamadaka, and M. B. Rao.

This article is part of the topical collection "Celebrating the Centenary of Professor C. R. Rao" guest edited by Ravi Khattree, Sreenivasa Rao Jammalamadaka, and M. B. Rao.

✉ S. Rao Jammalamadaka
rao@pstat.ucsb.edu

Brian Wainwright
wainwright@pstat.ucsb.edu

Qianyu Jin
qianyu.homu@gmail.com

[1] Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

[2] JPMorgan Chase & Co., New York, USA

Published online: 10 March 2021

1 Introduction

This paper provides a method for clustering density curves around the unit circle by approximating each such curve as a mixture of an appropriate number of von Mises (vM) distributions. The vM distribution, also known as the Circular Normal (CN) distribution, is one of the most commonly used models for directions in two-dimensions called circular data and has the density function,

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

where $0 \le \mu < 2\pi$ and $\kappa > 0$ are the mean direction and concentration parameter, respectively, and $I_p(\kappa)$ is the modified Bessel function of the first kind and order p. Although this distribution is unimodal, an appropriate mixture of such distributions can be used to model curves with multiple peaks. Indeed, similar to a corresponding result on the real line which says that any probability distribution on the real line can be approximated by a countable mixture of Gaussians (see [2, 10]), one can use a countable mixture of vM distributions to approximate any probability distribution on the circle.

Given that each curve can be approximated by such a mixture, the task of clustering these curves around the circle depends on defining an appropriate "distance" or a "divergence measure" between any two curves. To that end, we first consider such a distance or divergence measure between any two vM models in the next section, and then extend such measures to distances between any two vM *mixtures* in the subsequent section.

2 Measures of Distance and Divergence Between Two vM Models

In this section, we describe two different measures to find the divergence between two such vM models, one an actual L^2 distance, and the other a measure based on symmetrized Kullback–Leibler divergence. Later, we show that either of these provides a good "distance" measure between curves around a circle, helping in clustering them, and that they indeed provide very similar results. We start with the following useful result.

Lemma 1 *For any two vM distributions $f \sim vM(\mu_1,\kappa_1)$ and $g \sim vM(\mu_2,\kappa_2)$,*

$$\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha = \frac{I_0(\kappa)}{2\pi I_0(\kappa_1)I_0(\kappa_2)}, \tag{1}$$

where

$$\kappa = \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2\cos(\mu_1-\mu_2)}.$$

Proof We have

$$\begin{aligned}
&\int_0^{2\pi} f(\alpha)g(\alpha)d\alpha \\
&= \frac{1}{4\pi^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{\kappa_1\cos(\alpha-\mu_1)+\kappa_2\cos(\alpha-\mu_2)}d\alpha \\
&= \frac{1}{4\pi^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} e^{(\kappa_1\cos\mu_1+\kappa_2\cos\mu_2)\cos\alpha+(\kappa_1\sin\mu_1+\kappa_2\sin\mu_2)\sin\alpha}d\alpha \\
&= \frac{I_0(\kappa)}{2\pi I_0(\kappa_1)I_0(\kappa_2)},
\end{aligned}$$

where

$$\begin{aligned}
\kappa &= \sqrt{(\kappa_1\cos\mu_1+\kappa_2\cos\mu_2)^2+(\kappa_1\sin\mu_1+\kappa_2\sin\mu_2)^2} \\
&= \sqrt{\kappa_1^2+\kappa_2^2+2\kappa_1\kappa_2\cos(\mu_1-\mu_2)}.
\end{aligned}$$

□

Note that this Lemma 1 is related to Result J on the convolution of two vM distributions, given in §2.2 of Jammalamadaka and SenGupta (2001) [9], and can also be obtained by replacing $(\theta-\mu_2)$ there by μ_2.

2.1 L^2 Distance Between Two vM Models

Using Lemma 1, the L^2 distance between any two vM distributions can be obtained and is given by the following

Proposition 1 *For any two vM distributions $f \sim vM(\mu_1,\kappa_1)$ and $g \sim vM(\mu_2,\kappa_2)$, the L^2 distance is*

$$L_2(f,g) = \frac{1}{2\pi}\left(\frac{I_0(2\kappa_1)}{I_0(\kappa_1)^2}+\frac{I_0(2\kappa_2)}{I_0(\kappa_2)^2}-\frac{2I_0(\kappa)}{I_0(\kappa_1)I_0(\kappa_2)}\right), \tag{2}$$

where

$$\kappa = \sqrt{\kappa_1^2+\kappa_2^2+2\kappa_1\kappa_2\cos(\mu_1-\mu_2)}.$$

Proof By Lemma 1,

$$\begin{aligned}
\int_0^{2\pi} f^2(\alpha)\,d\alpha &= \frac{I_0(\sqrt{\kappa_1^2+\kappa_1^2+2\kappa_1\kappa_1\cos(\mu_1-\mu_1)})}{2\pi I_0(\kappa_1)I_0(\kappa_1)} \\
&= \frac{I_0(2\kappa_1)}{2\pi(I_0(\kappa_1))^2},
\end{aligned}$$

and similarly

$$\int_0^{2\pi} g^2(\alpha)\, d\alpha = \frac{I_0(2\kappa_2)}{2\pi(I_0(\kappa_2))^2}.$$

Thus,

$$\begin{aligned} L_2(f,g) &= \int_0^{2\pi} (f(\alpha) - g(\alpha))^2\, d\alpha \\ &= \int_0^{2\pi} f^2(\alpha)\, d\alpha + \int_0^{2\pi} g^2(\alpha)\, d\alpha - 2\int_0^{2\pi} f(\alpha)g(\alpha)\, d\alpha \\ &= \frac{1}{2\pi}\left(\frac{I_0(2\kappa_1)}{(I_0(\kappa_1))^2} + \frac{I_0(2\kappa_2)}{(I_0(\kappa_2))^2} - \frac{2I_0(\kappa)}{I_0(\kappa_1)I_0(\kappa_2)}\right), \end{aligned}$$

giving us the desired result. □

2.2 Kullback–Leibler (KL) Divergence Between Two vM Models and Its Symmetric Version

2.2.1 Kullback–Leibler (KL) Divergence Between Two vM Models

We first obtain a closed form solution to the standard KL divergence between any two vM models.

Proposition 2 *For any two vM distributions $f \sim vM(\mu_1, \kappa_1)$ and $g \sim vM(\mu_2, \kappa_2)$, the Kullback–Leibler divergence measure is given by*

$$\mathrm{KL}(f,g) = \log I_0(\kappa_2) - \log I_0(\kappa_1) + \kappa_1 A(\kappa_1) - \kappa_2 \cos(\mu_1 - \mu_2) A(\kappa_1), \tag{3}$$

where $A(\kappa)$ is the ratio of modified Bessel functions of the first kind, given by $A(\kappa) = I_1(\kappa)/I_0(\kappa)$.

Proof Let $f(\alpha;\mu_1,\kappa_1)$ and $g(\alpha;\mu_2,\kappa_2)$ be two distinct vM distributions, for $0 \le \alpha < 2\pi$. Then, the KL divergence between f and g is given by

$$
\begin{aligned}
\mathrm{KL}(f,g) &= \int_0^{2\pi} \log\left[\frac{f(\alpha)}{g(\alpha)}\right] f(\alpha)\, d\alpha \\
&= \int_0^{2\pi} \left\{\log\left[I_0(\kappa_2)\right] - \log\left[I_0(\kappa_1)\right] + \kappa_1 \cos(\alpha-\mu_1) - \kappa_2 \cos(\alpha-\mu_2)\right\} \\
&\quad \cdot \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha \\
&= \log\left[I_0(\kappa_2)\right] - \log\left[I_0(\kappa_1)\right] + \kappa_1 \frac{I_1(\kappa_1)}{I_0(\kappa_1)} \\
&\quad - \kappa_2 \int_0^{2\pi} \cos(\alpha-\mu_2) \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha.
\end{aligned}
$$

Writing the integrand in the last term as

$$
\begin{aligned}
\cos(\alpha-\mu_2) &= \cos(\alpha-\mu_1+\mu_1-\mu_2) \\
&= \cos(\alpha-\mu_1)\cos(\mu_1-\mu_2) - \sin(\alpha-\mu_1)\sin(\mu_1-\mu_2),
\end{aligned}
$$

and using the relations (see, e.g., Jammalamadaka and SenGupta (2001) [9], p.36) below,

$$
\begin{aligned}
\frac{1}{2\pi}\int_0^{2\pi} \cos(p\theta)\exp(\kappa\cos\theta)\, d\theta &= I_p(\kappa), \\
\frac{1}{2\pi}\int_0^{2\pi} \sin(n\theta)\exp(\kappa\cos\theta)\, d\theta &= 0,
\end{aligned}
$$

the last term becomes

$$
\begin{aligned}
&\kappa_2 \int_0^{2\pi} \cos(\alpha-\mu_2) \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha \\
&= \kappa_2 \int_0^{2\pi} \left[\cos(\mu_1-\mu_2)\cos(\alpha-\mu_1) - \sin(\mu_1-\mu_2)\sin(\alpha-\mu_1)\right] \\
&\quad \cdot \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha \\
&= \kappa_2 \cos(\mu_1-\mu_2) \int_0^{2\pi} \cos(\alpha-\mu_1) \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha \\
&\quad - \kappa_2 \sin(\mu_1-\mu_2) \int_0^{2\pi} \sin(\alpha-\mu_1) \frac{e^{\kappa_1 \cos(\alpha-\mu_1)}}{2\pi I_0(\kappa_1)}\, d\alpha \\
&= \kappa_2 \cos(\mu_1-\mu_2) \frac{I_1(\kappa_1)}{I_0(\kappa_1)},
\end{aligned}
$$

giving us the desired result. □

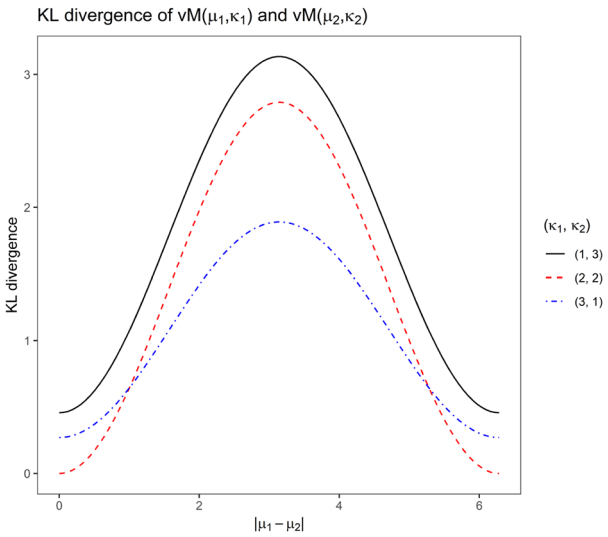

Fig. 1 KL-Div. vs $|\mu_1 - \mu_2|$, for fixed κ_1, κ_2

Remark 1 In general, such a KL divergence is not symmetric, but in the special case when $\kappa_1 = \kappa_2$ say with a common concentration κ, the D_{KL} divergence between two vM distributions becomes symmetric, and we have

$$\mathrm{KL}(f, g) = \mathrm{KL}(g, f) = \kappa A(\kappa)[1 - \cos(\mu_1 - \mu_2)],$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$.

Graphical evidence for the lack of symmetry for KL divergence can be seen in Fig. 1 where the black solid like representing the KL divergence for $\kappa_1 = 1$, $\kappa_2 = 3$ is clearly not identical to the KL divergence when $\kappa_1 = 3$ and $\kappa_2 = 1$, denoted by the dashed blue line.

2.2.2 A Symmetrized Kullback–Leibler (SKL) Distance Between Two vM Models

As remarked above, although this KL divergence becomes symmetric in the special case when the concentration parameters are the same, in general it is well known the KL measure lacks symmetry. We now consider a simple symmetric version of the KL divergence namely

$$\mathrm{SKL}(f, g) = D_{\mathrm{KL}}(f, g) + D_{\mathrm{KL}}(g, f).$$

We will refer to this as the symmetric KL divergence or SKL. From the preceding Proposition 2, it is easy to check the following conclusion.

Proposition 3 *The symmetric KL divergence SKL, between two vM distributions* $f \sim vM(\mu_1, \kappa_1)$ *and* $g \sim vM(\mu_2, \kappa_2)$ *is given by*

$$\mathrm{SKL}(f,g) = \kappa_1 A(\kappa_1) + \kappa_2 A(\kappa_2) - \cos(\mu_1 - \mu_2)\big[\kappa_2 A(\kappa_1) + \kappa_1 A(\kappa_2))\big]$$

where $A(\kappa)$ *is the ratio of modified Bessel functions of the first kind, given by* $A(\kappa) = I_1(\kappa)/I_0(\kappa)$.

Remark 2 Jensen–Shannon divergence: One may also consider a slightly more general symmetrized version, called the Jensen–Shannon divergence given by

$$\mathrm{SKL}_{\mathrm{JS}}(f,g) = \frac{D_{\mathrm{KL}}\left(f, \frac{f+g}{2}\right) + D_{\mathrm{KL}}\left(g, \frac{f+g}{2}\right)}{2}$$

noting that mixtures of vM distributions are known to be identifiable (see, e.g., Frazer et al. [3] and Holzmann et al. [5]).

2.3 A Visual Comparison of the L^2 and SKL Measures

Figures 2 and 3 illustrate these two divergence measures side-by-side in terms of how they change by varying $|\mu_1 - \mu_2|$, or $\kappa_2 - \kappa_1$, or both. Although the magnitudes of the two measurements are different—indeed SKL is ~ 10 times the L^2 distance given the same parameters in vM mixture—they provide comparable results judging by the similarities in the shapes of the line plots and 3D plots between the two.

3 L^2 Distance and the SKL Divergence for vM Mixtures

As suggested earlier, we plan to use vM mixtures with an appropriate number of components to model any curve around the circle. Consider any 2 such mixtures h_1 and h_2 denoted by

$$h_1(\alpha) = \sum_{i=1}^{k} p_i f_i(\alpha) \qquad \text{and} \qquad h_2(\alpha) = \sum_{j=1}^{l} q_j g_j(\alpha),$$

with the usual restrictions on the mixture proportions, $\sum_i p_i = \sum_j q_j = 1$. Then, the L^2 distance between the two mixtures h_1 and h_2 can easily be defined using Proposition 1 above, and we have

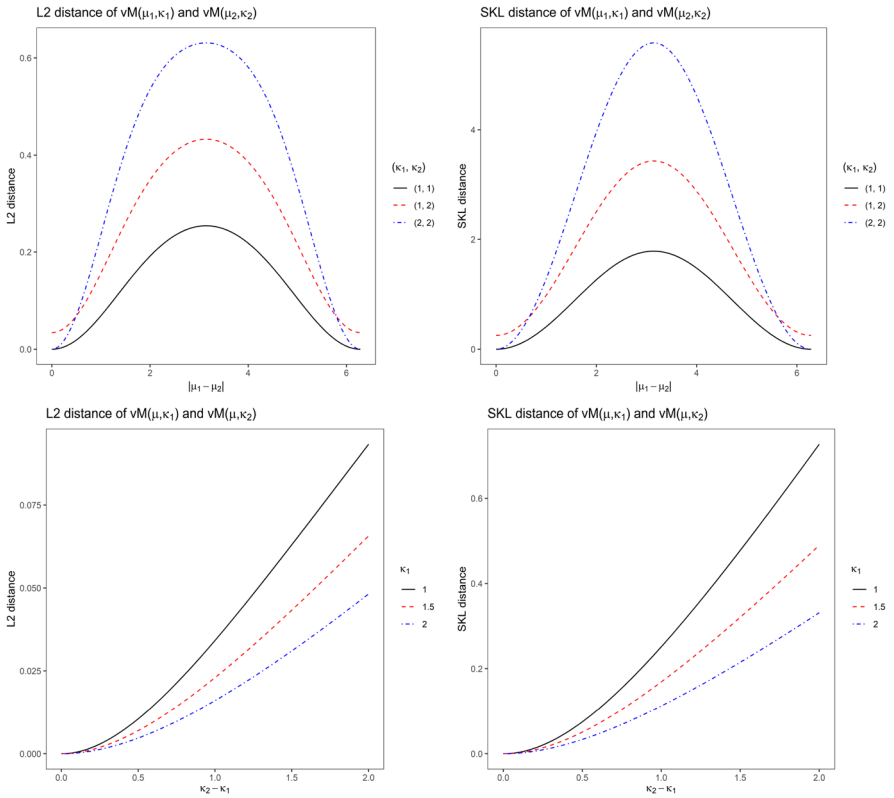

Fig. 2 From left to right: L^2 distance and SKL distance vs $|\mu_1 - \mu_2|$, for fixed κ_1, κ_2; L^2 distance and SKL distance vs $\kappa_2 - \kappa_1$, for fixed $\mu_1 = \mu_2 = \mu$ and κ_1. Note that the y-axis scales are different for L^2 and SKL distance

$$
\begin{aligned}
L_2(h_1, h_2) &= \int_0^{2\pi} \left[\sum_{i=1}^{k} p_i f_i(\alpha) - \sum_{j=1}^{l} q_j g_j(\alpha) \right]^2 d\alpha \\
&= \sum_{i=1}^{k} \sum_{i'=1}^{k} p_i p_{i'} \int_0^{2\pi} f_i(\alpha) f_{i'}(\alpha)\, d\alpha \\
&\quad + \sum_{j=1}^{l} \sum_{j'=1}^{l} q_j q_{j'} \int_0^{2\pi} g_j(\alpha) g_{j'}(\alpha) d\alpha \\
&\quad - 2 \sum_{i=1}^{k} \sum_{j=1}^{l} p_i q_j \int_0^{2\pi} f_i(\alpha) g_j(\alpha)\, d\alpha.
\end{aligned}
$$

When extending the idea of SKL divergence to mixtures of vM distributions, we must proceed carefully as no closed form expression exists for

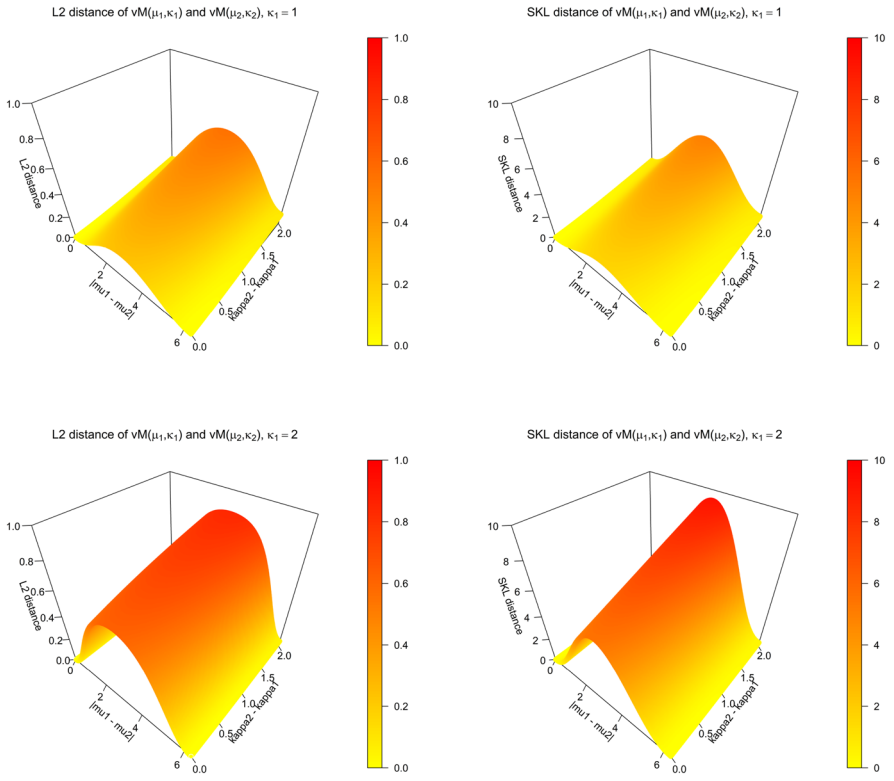

Fig. 3 3D plots for L^2 distance (left) and SKL distance (right) vs $|\mu_1 - \mu_2|$ and $\kappa_2 - \kappa_1$ for fixed κ_1. Note that the color scales are different for L^2 and SKL distance

$$D_{\mathrm{KL}}(h_1, h_2) = \int_0^{2\pi} \log\left[\frac{h_1(\alpha)}{h_2(\alpha)}\right] h_1(\alpha)\, d\alpha,$$

unlike what we had for the single components. Because of this, we need to resort to numerical integration techniques to evaluate this divergence.

4 A Simulation Study to Judge the Effectiveness of Clustering via L^2 and SKL Measures

Our simulation procedure starts with given vM mixtures with known numbers of components and pre-specified parameter values for each component. Our goal is to assess how these two measures perform in identifying the known clusters, under varying distributional conditions. The cases we choose and the parameter values selected, are given in Table 1 below.

Each scenario consists of 3 vM mixture distributions—two 2-component mixtures and one 3-component mixture—with different parameter values. Case 1

Table 1 Simulated Data Parameter Values

Case	Mixture	Component	μ (rad.)	κ	p
Case 1	vM-mix A	$f_1(\alpha;\mu,\kappa)$	0	4.0	0.5
		$f_2(\alpha;\mu,\kappa)$	$2\pi/3$ (≈ 2.09)	4.0	0.5
	vM-mix B	$f_1(\alpha;\mu,\kappa)$	π (≈ 3.14)	3.0	0.5
		$f_2(\alpha;\mu,\kappa)$	$5\pi/3$ (≈ 5.24)	3.0	0.5
	vM-mix C	$f_1(\alpha;\mu,\kappa)$	0	5.0	0.333
		$f_2(\alpha;\mu,\kappa)$	$2\pi/3$ (≈ 2.09)	5.0	0.333
		$f_3(\alpha;\mu,\kappa)$	$4\pi/3$ (≈ 4.19)	5.0	0.333
Case 2	vM-mix A	$f_1(\alpha;\mu,\kappa)$	0	4.0	0.75
		$f_2(\alpha;\mu,\kappa)$	$2\pi/3$ (≈ 2.09)	4.0	0.25
	vM-mix B	$f_1(\alpha;\mu,\kappa)$	π (≈ 3.14)	3.0	0.25
		$f_2(\alpha;\mu,\kappa)$	$5\pi/3$ (≈ 5.24)	3.0	0.75
	vM-mix C	$f_1(\alpha;\mu,\kappa)$	0	5.0	0.20
		$f_2(\alpha;\mu,\kappa)$	$2\pi/3$ (≈ 2.09)	5.0	0.60
		$f_3(\alpha;\mu,\kappa)$	$4\pi/3$ (≈ 4.19)	5.0	0.20

contains vM mixtures that have equal mixture weights p, moderately high κ's, and with differences between μ's that are large enough such that the number of components k equals the number of modes. Case 2 has unequal p's and thus unequal heights for each mode. These cases aim at showing how sample-based parameter estimation and component selection results are affected by the underlying mixture distributions.

With the parameter values set for the mixture distribution, we take $n = 5$ samples, each sample being of size $m = 100$. We use the 100 observations from each sample to fit a mixture distribution, estimating the mixture model parameters via the EM algorithm proposed by Dhillon and Sra [8] and Banerjee et al. [1], and implemented by Hornik and Grün in the `R` statistical software package `movMF` [7][1]. Since the parameter estimation method does not automatically select the number of mixture components, we apply the BIC criterion to select the number of components, say between 2, 3, 4, or 5. From each of the 3 *true* mixture distributions for each case study, we generate $n = 5$ estimated mixture distributions using the samples drawn.

Figure 4 graphically demonstrates the efficacy of the estimated mixtures in recovering the true mixture. In the grid of plots, the rows represent the case study and the columns represent the three different vM mixture distributions within each case. In any particular plot, the black dashed line is the *true* density, while the colored solid lines are the five density estimates associated with the five samples. Case 1 shows good consistency between true and estimated density curves, except for the red curve in vM mixture 3 which is flat and gives only k = 1 component. Case 2

[1] This algorithm and implementation describe fitting the parameters of von Mises-Fisher (vMF) mixture models, a direct higher-dimensional extension of the vM distribution with observations on the unit sphere/hypersphere. In two dimensions, this reduces to the vM distribution.

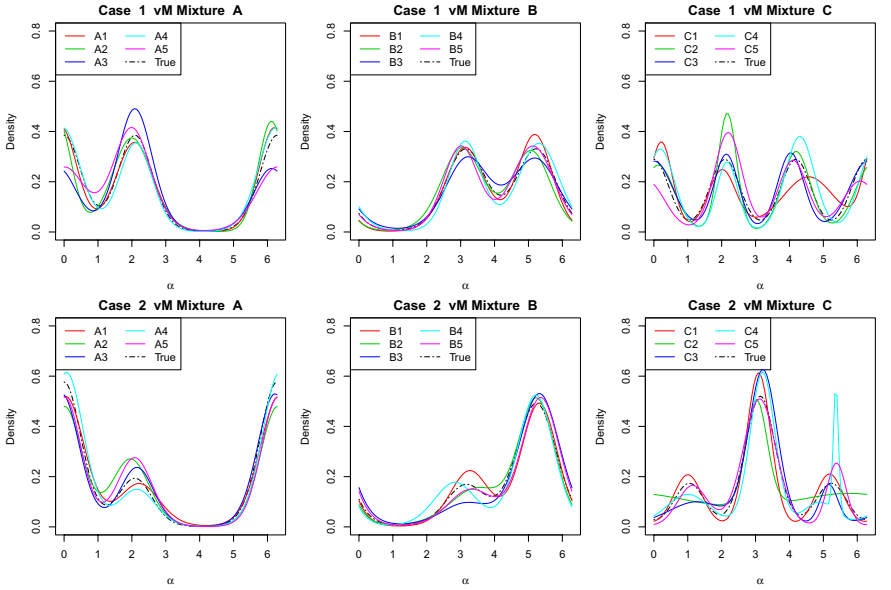

Fig. 4 Estimated Density Curves vs. True Density Curves. The three plots in each row correspond to the three vM mixtures in each case. Black dashed lines are the true density curves, and colored solid lines are estimated density curves generated from our five samples

shows that smaller κ values can impair this consistency, as wrong values of k are often chosen.

4.1 Clustering of Simulated Data with L^2 and SKL Measures

After estimating vM mixture parameters from simulated samples for each case study, we can compute L^2 or SKL distances between each pair of samples and obtain distance matrices. Then, a hierarchical clustering method is used to discover the clusters. We selected the *complete linkage* method and show the clustering results in heatmaps and dendrograms—heatmaps are to show the magnitudes of distance measures between any given pair of estimated vM mixtures, and dendrograms are to show the structures of clustering hierarchy. Ideally, what we would like to see are the clustering procedure, when stopped at 3 clusters, completely separates the estimated density curves based on which mixture distribution their samples are generated from, i.e., each of the 3 clusters should contain the 5 densities estimated from the common true density.

Both L^2-based clustering and SKL-based clustering are able to recover the correct cluster membership for all the samples as illustrated in Figs. 5 and 6, which display row end dendrograms and heatmaps for these two clustering metrics. For both Cases 1 and 2, within-cluster L^2 distances are very close to zero, while between-cluster L^2 distances are much larger than within-cluster L^2 distances. The

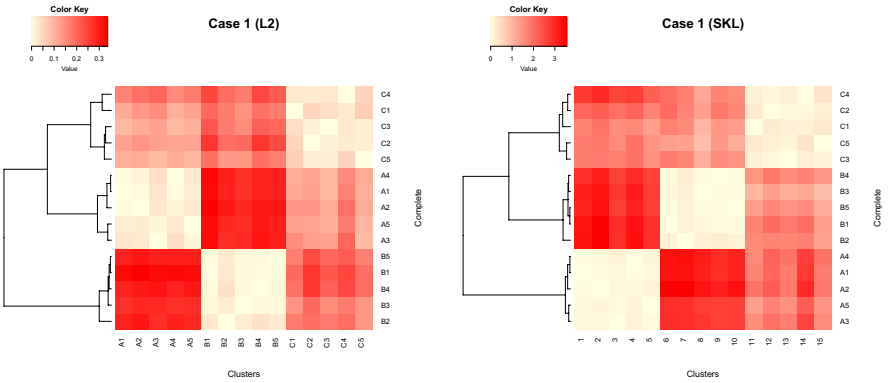

Fig. 5 L^2 Clustering Heatmap of Simulated Data. Color key represents the L^2 distance

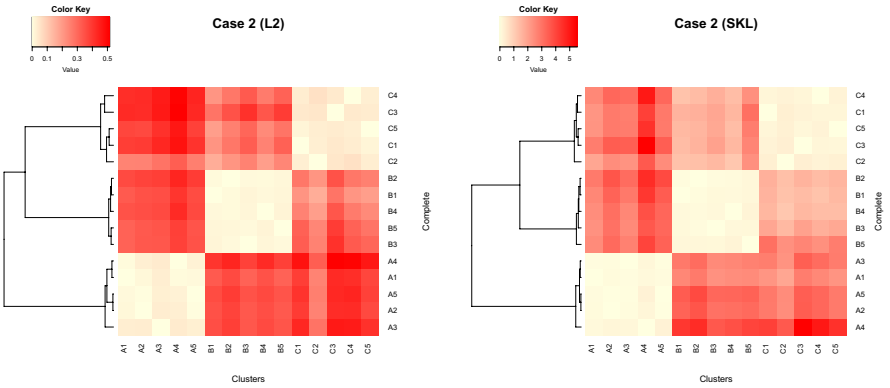

Fig. 6 SKL Clustering Heatmap of Simulated Data. Color key represents the SKL distance

Table 2 L_2 and SKL Hierarchy Agreement for Simulation Study

Euclidean	Cosine	Cophenetic	Gamma
0.0387	0.9258	0.7320	0.7410

heatmap patterns suggest that, for arbitrary samples denoted by s_1, s_2, t_1 and t_2, the distance $L_2(s_1, t_1) \approx L_2(s_2, t_2)$ as long as s_1, s_2 are in the same cluster and t_1, t_2 are in the same cluster. Similar behaviors are observed for SKL distances.

4.2 Agreement Between L^2 and SKL Clustering Hierarchy

Figures 5 and 6 helped to visually establish the similarities between the hierarchies of L^2 and SKL clustering through the resemblance of heatmap patterns and identical cluster membership when the number of clusters is fixed at 3. In addition, we will

also quantify the similarities by computing the hierarchy agreement of their respective ultrametrics in Table 2.

For the collection of subsets Ω induced by a hierarchical clustering scheme, the distance h_{ij} between curves i and j is defined as the height of the node that generates the smallest set containing both elements, such that the collection of heights, $\{h_{ij}; \text{for } i, j \in \Omega\}$, satisfies the ultrametric inequality,

$$h_{ij} \leq max\big(h_{ik}, h_{jk}\big) \quad \forall i, j, k \in \Omega \quad [3]$$

There are a variety of metrics that can be employed to compare the respective hierarchy ultrametrics, but we will restrict our analysis to looking at four of the metrics that come standard in `R` package `clue` [6]:

Euclidean: when we take d as the square root of the sum of the squared differences of ultrametrics, the Euclidean agreement is given as $1/(1+d)$.
Cophenetic: product-moment correlation of the respective ultrametrics, also known as the cophenetic correlation coefficient.
Cosine: the cosine of the angle between the respective ultrametrics.
Gamma: a linear transformation of Kruskal's gamma, $1-d$, where we take d as the rate of inversions between respective ultrametrics, $u_{ij} < u_{kl}$ and $v_{ij} > v_{kl}$, for pairs (i, j) and (k, l).

For a fuller treatment of ultrametrics in hierarchical comparisons, see [4].

For all of these four metrics, the value falls between 0 and 1, and a value closer to 1 indicates higher similarity between the ultrametrics of the two clustering hierarchies. While the Euclidean metric is very small, the other three metrics, in particular the cosine metric, are all fairly close to 1, which indicates a high level of similarity.

5 A Real Data Application—Eye Imaging Analysis

5.1 Optical Coherence Tomography

Purely for illustrative purposes as to how these clustering techniques work, we now look at a small sample of 100 curves based on eye imaging data. A comprehensive analysis of the original data set, from which this small sample is selected for illustration, is in preparation and will appear elsewhere. Optical Coherence Tomography (OCT) is a noninvasive imaging technique that uses a broadband light source partitioned into a reference beam and sample beam to generate a reflectivity versus depth profile that details an approximate in-vivo retinal biopsy. The top panel of Fig. 7 shows a diagram of the retina, with the optic disc, optic cup, as well as labels of the four quadrants of the eye, superior, nasal, inferior, and temporal.

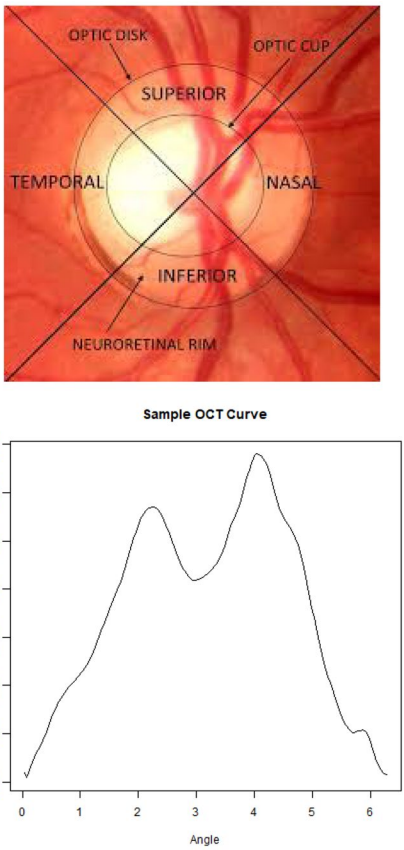

Fig. 7 From top to bottom, A scanned image of the eye, and an example OCT data plotted on the line

5.2 General Characteristics of the Illustrative Data

Although an OCT measurement for each eye can be taken continuously around the eye, for our purposes, it is measured on a discrete grid of regularly spaced angles around the eye. Since the measurements are taken radially around the circumference of the retina, these curves are circular in nature and are amenable to methods described here. The bottom panel in Fig. 7 is an example of a randomly selected OCT curve plotted on a linear scale. The task is to cluster these circular density curves into an unknown number K homogeneous groups based on curve features, via the estimated vM mixtures.

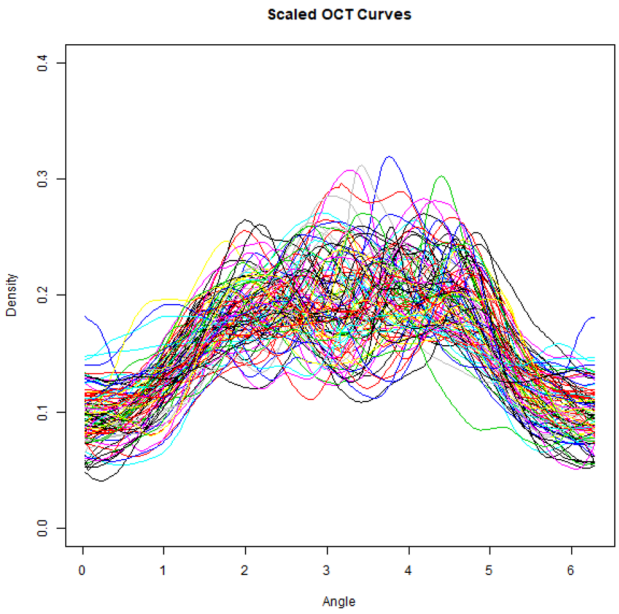

Fig. 8 A sample of 100 scaled OCT Curves

We scaled these curves so that each represents an empirical circular density, and our proposed density clustering methodology is applicable. All of the 100 scaled curves are shown in Fig. 8. The results of hierarchical clustering by using L^2 and SKL distances are represented by the dendrograms in Fig. 9.

Because this task is unsupervised, without a ground truth with respect to the number of clusters or cluster membership, we will look at cluster agreement between the two schemes via their respective ultrametrics in Table 3. Euclidean metric is very small, Cosine metric is close to 1, and the rest two metrics are in the middle. This implies that L^2 and SKL clustering hierarchies are fairly similar.

6 Concluding Remarks

The principal aim of this paper is to describe methods of functional clustering, when dealing with periodic curves around a circle, in particular circular density curves. It is shown that this can be accomplished successfully by approximating each density curve by a vM mixture, and clustering them using the distance matrix computed from vM mixture parameters. After demonstrating via simulations that this technique works successfully, it is applied on a practical OCT data set. The **R**-code for fitting such mixtures and clustering them, is available by requesting the authors.

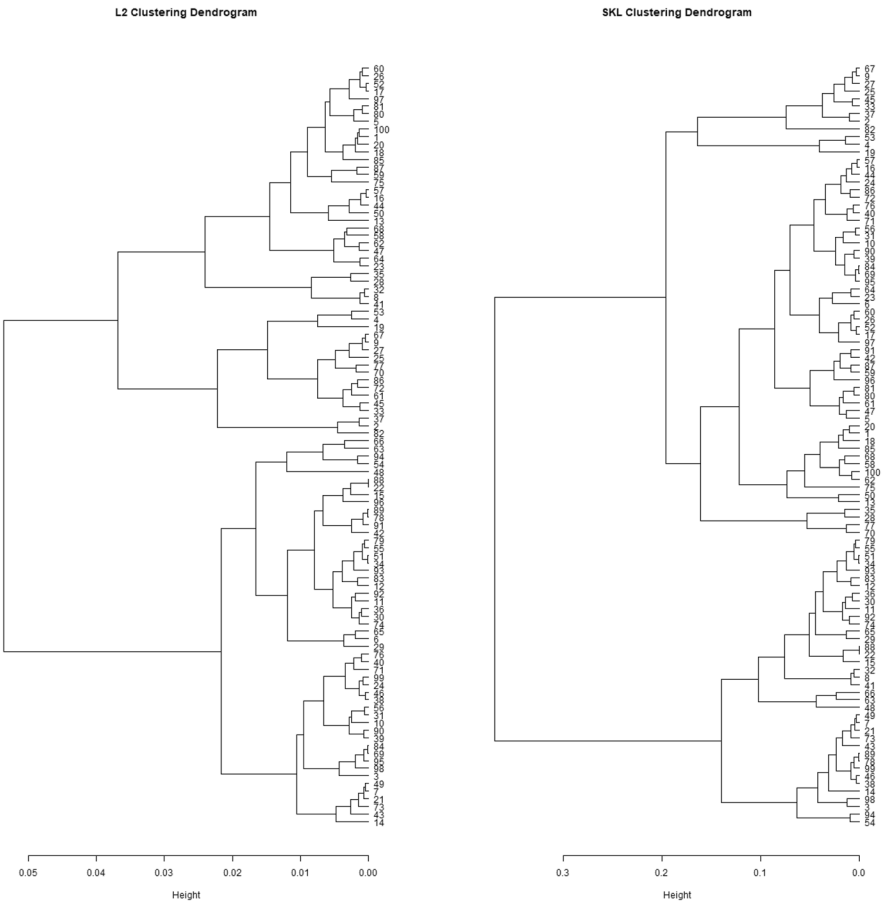

Fig. 9 Dendrograms for L^2 (left) and SKL (right) Hierarchical Clustering

Table 3 L_2 and SKL Hierarchy Agreement for OCT Data

Euclidean	Cosine	Cophenetic	Gamma
0.0565	0.8658	0.3887	0.6678

References

1. Banerjee A et al (2005) Clustering on the unit hypersphere using von Mises- Fisher distributions. J Mach Learn Res 6:1345–1382
2. Ferguson TS (1983) Bayesian Density Estimation by Mixtures of Normal Distributions. In: Rizvi M, Rustagi J, Siegmund D (eds) Recent Advances in Statistics, vol 24. Academic Press, Cambridge, pp 287–302
3. Fraser MD, Hsu YS, Walker JJ (1981) Identifiability of finite mixtures of Von Mises distributions. Ann Stat 9(5):1130–1131
4. Gordon AD (1999) Classification, 2nd edn. Monographs on Statistics& Applied Probability. CRC Press, Chapman & Hall/CRC, Boca Raton

5. Holzmann H, Munk A, Stratmann B (2004) Identifiability of finite mixtures-with applications to circular distributions. Sankhyā 66(3):440–449
6. Hornik Kurt (2019) clue: Cluster ensembles. R package version 0.3-57. https://CRAN.R-project.org/package=clue
7. Hornik Kurt, Grün Bettina (2014) "On maximum likelihood estimation of the concentration parameter of von Mises-Fisher distributions". In: Computational Statistics 29.5, pp. 945-957. ISSN: 1613-9658. https://doi.org/10.1007/s00180-013-0471-0
8. Dhillon IS, Sra S (2003) Modeling Data using Directional Distributions. Tech. rep. TR-03-06. Austin, TX 78712: Department of Computer Sciences, The University of Texas at Austin, Jan. https://www.cs.utexas.edu/users/inderjit/public_papers/tr03-06.pdf
9. Jammalamadaka S.Rao, SenGupta A (2001) Topics In Circular Statistics. Vol. 5. Series on Multivariate Analysis. Singapore, London, New Jersey, Hong Kong: World Scientific Publishing Co. Pte. Ltd., ISBN: 9810237782
10. Teicher H (1960) "On the Mixture of Distributions". In: The Annals of Mathe- matical Statistics 31.1, pp. 55–73. http://www.jstor.org/stable/2237493