https://doi.org/10.56093/JISAS.V80I1.21
Available online at https://isas.live/newsite
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 80(1) 2026 249–260

Hierarchical Clustering on the Torus

Ashis SenGupta[1] and Moumita Roy[2]

[1]Department of Population Health Sciences, MCG, Augusta University, GA, USA and Indian Statistical Institute, Kolkata
[2]Midnapore College (Autonomous), Midnapore

Received 02 May 026; Revised 06 May 2026; Accepted 07 May 2026

SUMMARY

The aim of this paper is to introduce hierarchical clustering for bivariate circular or toroidal data. Here a mixture model approach is used as for model based clustering. Here, the clustering is performed in two stages, the first stage being based on the marginal distribution of one of the circular variables and the second stage being based on the conditional distribution of the other variable given the former. In particular, two types of such mixture models are constructed. A real life application on gene data is made to illustrate the use of the proposed approach. Comparison of the two models is also done based on this example.

Keywords: Akaike information criterion (AIC); Bayesian information criterion (BIC); Bivariate von-Mises distribution; Expectation-Maximization algorithm; Hierrachical; Clustering.

1. INTRODUCTION

In many emerging real-life situations, it is quite common to encounter data on possibly two dependent or bivariate circular random variables, i.e data on the surface of the torus S^1 x S^1.Popular examples of such data on the torus include wind directions measured at two time points of the same day and at the same location over the four seasons, gene expression microarray of circadian-related genes in two tissues, etc. Not much literature is available for distance based or model based clustering for such data. Although recently, [12] has introduced model based clustering methods for clustering toroidal data. The method is illustrated shortly in Section 4. In this paper, our aim is to further develop other model-based clustering methods for toroidal data. Here we introduce a hierarchical method for clustering toroidal data in Section 5. This method is model based clustering and is mainly performed in two stages. The first stage involves the formation of the initial clusters based on the mixture model of the marginal distribution of one circular variable. After the formation of the initial clusters, the next stage involves the formation of sub clusters of each of the initial clusters obtained. This stage uses the conditional distribution of the other circular variable given the former. Two mixture models are adopted here. In Sub section 5.1, the process of estimation of the parameters of the distributions under Model 1 and Model 2 for the first stage clustering are discussed. In the first stage of Model 1, the marginal distribution considered is that of the popular von-Mises distribution. The first stage of Model 2 is based on the marginal distribution of the sine version of the popular bivariate von-Mises distribution. The process of estimation of parameters for the second stage clustering is described in Subsection 5.2, where for Mode 1, the conditional distribution of the other variable given the previous is taken as another von-Mises distribution and Model 2 is based on the conditional distribution of the same bivariate distribution as taken for the first stage. In Section 6, the technique of allocation of observations to

Corresponding author: Ashis SenGupta
E-mail address: amsseng@gmail.com

the clusters is described. The methods for determining the optimum number of clusters is given in Section 7. A real life application on gene data is made to illustrate the use of the proposed approaches in Section 8. The comparison of the performance of the clustering under the two models is illustrated in Section 9. Finally, some concluding remarks are presented in Section 10 and some acknowledgement is given in Section 11.

2. MODEL BASED CLUSTERING ON THE TORUS

A mixture model approach based on the joint distribution of the two circular variables was introduced in [12] for model-based clustering of toroidal data. We considered two possible situations. In the first situation, we considered the marginal distribution of one of the circular variables to be the popular univariate von-Mises distribution. The joint distribution for the component in the mixture model was then formulated using the conditional distribution of the other circular variable given the first. Next we considered the situation where we adopted the popular bivariate von-Mises distribution, as both its conditionals (but not the marginals) are the popular von-Mises distributions. Expectation-Maximization (EM) algorithm is used for estimating the various parameters of the mixture distributions. The convergence of the EM algorithm was ensured for the mixture models used as proved in [15] and also later extended in [8]. On convergence of the algorithm, the observations are allocated to the clusters according to the maximum conditional probability of membership. The methods for the optimum number of clusters is determined using the usual AIC and BIC criterion. It appeared from the clustering plot of the optimum number of clusters that there are more number of clusters due to the positioning of the two circular variables on the planar scale. But using the fact that the points on the left and right side of the X-axis for one circular variable and the top and the bottom edges of the Y-axis of the other variable are within close proximities, it was quite clear that actually the optimum number of clusters only is obtained in the plot. A measure for homogeneity of clusters based on a notion of distance for the toroidal topology was introduced. Smaller the distance, the better was the cluster homogeneity defined. A real life application on data of genes of heart and liver tissues was made of the above proposed approaches. The objective of this clustering could be helpful for the doctors to identify the genes related to heart and liver tissues which are similarly affected so that they can accordingly conduct the required proper treatment. Finally, a comparison of the performance of the two models was also done. The optimum number of clusters obtained under both the models was 4. Model 2 performed better than Model 1 for the clustering of the above data both in terms of the usual BIC and AIC criterion. The later model also performed better in terms of cluster homogeneity as it yielded smaller average toroidal distance measure as introduced in [12].

3. HIERARCHICAL CLUSTERING

A hierarchical method for clustering toroidal data is introduced here. This method is model based clustering and is mainly performed in two stages. The first stage involves the formation of the initial clusters based on the mixture model of the marginal distribution of one circular variable. After the formation of the initial clusters, the next stage involves the formation of sub clusters of each of the initial clusters obtained. This stage uses the conditional distribution of the other circular variable given the former. Two mixture models are adopted here.

3.1 First stage clustering

The first stage clustering is based on the mixture model of the marginal distribution of one of the circular variables say, θ.

3.1.1 Model 1

Suppose the marginal distribution of θ_i is specified, e.g. as a von-Mises distribution with mean direction μ and concentration parameter κ. Then the distribution of θ_i is given by

$$f(\theta_i) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos\{\theta_i - \mu\}\}] \quad (1)$$

The first stage clustering is based on the mixture model

$$f(\theta;\psi) = \sum_{j=1}^{p} \pi_j f_j(\theta_i;\eta_j) \quad (2)$$

where p denotes the number of mixture components,

$\eta = (\mu_1,...,\mu_p,\kappa)'$, $\psi = (\eta', \pi_1,...,\pi_{p-1})$ $\left(\text{since } \pi_p = 1 - \sum_{j=1}^{p-1} \pi_j\right)$ denotes the parameter vectors of the mixture model, $\eta_j = (\mu_j,\kappa)'$, $(j = 1,...,p)$ denotes the parameter vector of the j^{th} component of the mixture model and

$$f_j(\theta_i;\eta_j) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa\cos\{\theta_i - \mu_j\}] \quad (3)$$

3.1.2 Estimation of parameters in Model 1

We use EM algorithm to estimate the parameters of the mixture model. Let z_{ij} be an indicator variable

$$z_{ij} = \begin{cases} 1 & \text{if } \theta_i \text{ comes from the } j^{th} \text{ component of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

Using the complete data vector $(z_{ij},\theta_i)'$, the complete data log-likelihood function for ψ is given by

$$L_c(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij} \ln f_j(\theta;\eta_j)$$

$$= -n \ln 2\pi - n \ln I_0(\kappa) + \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij}\left[\kappa\cos\{\theta_i - \mu_j\}\right] \quad (4)$$

If the z_{ij}'s are observable, then the MLE of π_j is simply given by $\hat{\pi}_j = \sum_{i=1}^{n} z_{ij}/n$.

The likelihood equations for estimating the parameters $\mu_j,(j = 1,...,p)$, $\kappa)$ respectively are given as

$$\sum_{i=1}^{n} \hat{z}_{ij}\left[\hat{\kappa}\sin\{\theta_i - \hat{\mu}_j\}\right] = 0 \quad (5)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{p} \hat{z}_{ij}\cos\{\theta_i - \hat{\mu}_j\} - \frac{nI_1(\hat{\kappa})}{I_0(\hat{\kappa})} = 0, \left(I_1(\kappa) = \frac{d}{d\kappa} I_0(\kappa)\right) \quad (6)$$

Hence solving the likelihood equations (5) and (6), the maximum likelihood estimators of the parameters are given by

$\hat{\mu}_j = \arctan^*(\bar{\theta}_1) \qquad j = 1,...,p$

where

$$\bar{\theta}_1 = \frac{\sum_{i=1}^{n} \hat{z}_{ij}\sin\theta_i}{\sum_{i=1}^{n} \hat{z}_{ij}\cos\theta_i} \quad (7)$$

and arctan* defines the quadrant specific angle as given in [3].

$\hat{\kappa} = A^{-1}(\bar{R}_1)$ where

$$\bar{R}_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} \hat{z}_{ij}\cos\{\theta_i - \hat{\mu}_j\}}{n} \text{ and } A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} \quad (8)$$

The steps of the EM algorithm are presented as follows:

Step 1: A random sample of size of the number of components p generated from uniform distribution over the range of the given data on the circular variable is taken as the initial estimate of the component mean direction μ. The mixing proportions π are initialized as a random sample of size p from a Dirichlet distribution with shape parameter 1. The initial estimate of the concentration parameter κ is generated from uniform distributions ranging from 0 to the value of the sample concentration of the circular variable.

Step 2: The marginal distribution used in Stage 1 that is the von-Mises density given by(1) is computed for the observed circular data θ_i using the initial values of the parameters for each of the p mixture components obtained in Step 1.

Step 3: The product of the mixing proportions π and the densities obtained in Step 2 are calculated for each component.

Step 4: The initial values of the indicator variable z_{ij} are obtained by dividing the product obtained for each j^{th} component in Step 3 by the sum of the products of all the components.

Step 5: In the first iteration k = 1, the mixing proportions π_j for each j^{th} component are estimated by the mean of the initial values of z_{ij} for the corresponding j^{th} component obtained in Step 4.

Step 6: In the first iteration k=1, the mean direction μ_j and concentration parameter κ are estimated by the solutions given by (7) and (8) substituting the initial values of z_{ij} obtained in Step 4.

Step 7: The log-likelihood equation given by (4) is calculated substituting the initial values of z_{ij}, the observed circular data θ_i and the estimates of the parameters obtained in Step 6.

Step 8: The Steps 2 to 7 are repeated until the difference between the values of the log-likelihood function of two consecutive iterations changes by an arbitrary small amount.

3.1.3 Model 2

Here the bivariate circular data (θ_i,ϕ_i), $i = 1,...,n$ is assumed to follow the sine model version of the bivariate von-Mises distribution whose probability density function (See [14]) is given by

$$f(\theta,\phi) = C.\exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - v) + \lambda \sin(\theta - \mu) \sin(\phi - v)\} \quad (9)$$

where C is the normalizing constant given by

$$C^{-1} = \int_0^{2\pi}\int_0^{2\pi} \exp\{\kappa_1\cos(\theta-\mu) + \kappa_2\cos(\phi-v) + \lambda\sin(\theta-\mu)\sin(\phi-v)\}\, d\theta d\phi \quad (10)$$

The marginal distribution of θ is given by

$$h(\theta) = C.2\pi I_0(\kappa)\exp\{\kappa_1\cos(\theta-\mu)\} \quad (11)$$

where $\kappa = \sqrt{\kappa_2^2 + \lambda^2\sin^2(\theta-\mu)}$.

Thus the first stage clustering is based on the mixture model

$$f(\theta;\psi) = \sum_{j=1}^{p} \pi_j f_j(\theta_i;\eta_j) \quad (12)$$

where p denotes the number of mixture components $\eta = (\mu_1,...,\mu_p,\kappa_1,\kappa_{i_j})$, $\psi = (\eta',\pi_1,...,\pi_{p-1})$ $\left(\text{since } \pi_p = 1-\sum_{j=1}^{p-1}\pi_j\right)$ denotes the parameter vectors of the mixture model, $\eta_j = (\mu_j,\kappa_1,\kappa_j)'$, $(j = 1,...,p,\ i = 1,...,n)$ denotes the parameter vector of the j^{th} component of the mixture model and

$$f_j(\theta_i;\eta_j) = C_j.2\pi I_0(\kappa_{i_j})\exp[\kappa_1\cos\{\theta_i-\mu_j\}] \quad (13)$$

3.1.4 Estimation of parameters in Model 2

We use EM algorithm to estimate the parameters of the mixture model. Let z_{ij} be an indicator variable

$$z_{ij} = \begin{cases} 1 & \text{if } \theta_i \text{ comes from the } j^{th} \text{ component of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

Using the complete data vector $(z_{ij},\theta_i)'$, the complete data log-likelihood function for ψ is given by

$$L_c(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij}\ln f_j(\theta;\eta_j)$$

$$= n\ln 2\pi + \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij}\left[\ln C_j + \ln I_0(\kappa_{i_j}) + \kappa_1\cos\{\theta_i-\mu_j\}\right] \quad (14)$$

If the z_{ij}'s are observable, then the MLE of π_j is simply given by $\hat{\pi}_j = \sum_{i=1}^{n} z_{ij}/n$.

The likelihood equations for estimating the parameters $\mu_j,(j=1,...,p)$ and κ_1 respectively are given as

$$\frac{\partial \ln C_j}{\partial \mu_j} + \sum_{i=1}^{n}\hat{z}_{ij}\left[\frac{\partial \ln I_0(\kappa_{i_j})}{\partial \mu_j} + \hat{\kappa}_1\sin(\theta_i-\mu_j)\right] = 0$$

$$\frac{\partial \ln C_j}{\partial \kappa_1} + \sum_{i=1}^{n}\sum_{j=1}^{p}\hat{z}_{ij}\left[\frac{\partial \ln I_0(\kappa_{i_j})}{\partial \kappa_1} + \cos(\theta_i-\hat{\mu}_j)\right] = 0$$

$\hat{z}_{ij}$ is the posterior probability that the i^{th} member of the sample with observed value θ_i belongs to the j^{th} component of the mixture.

We get the maximum likelihood estimates of the parameters by solving the above likelihood equations. However, we do not get any closed form expressions for the estimators from the likelihood equations. So, we use the "optim" function in R software to maximize the log-likelihood function given by (14) and provide suitable initial values to obtain the MLEs of the parameters.

The MLE of κ_{i_j} can be obtained by plugging in its expression, the estimates of μ_j and κ_1 obtained by solving the above equations.

The steps of the EM algorithm are presented as follows:

Step 1: Two random samples of size of the number of components p generated from uniform distribution over the range of the given data on the two circular variables are taken as the initial estimates of the component mean directions μ and v respectively. The mixing proportions π are initialized as a random sample of size p from a Dirichlet distribution with shape parameter 1. The initial estimates of the concentration parameters κ_1 and κ_2 are generated from uniform distributions ranging from 0 to the value of the sample concentration of the two circular variables respectively. The initial estimate of the interaction parameter λ is taken as the circular correlation between the two circular variables.

Step 2: The normalising constant for the marginal distribution in Stage 1 given by (10) are computed for each component of the mixture distribution.

Step 3: The initial value of the concentration parameter κ of the marginal distribution used in Stage 1 given by (11) is calculated from its relation with κ_2, λ and μ for each j^{th} component using their corresponding initial values obtained in Step 1.

Step 4: The marginal distribution used in Stage 1 given by (11) is computed for the observed circular data θ_i using the initial values of the parameters obtained in Steps 1 and 3 and the normalising constants obtained in Step 2 for each of the p mixture components .

Step 5: The product of the mixing proportions π and the densities obtained in Step 4 are calculated for each component.

Step 6: The initial values of the indicator variable z_{ij} are obtained by dividing the product obtained for each j^{th} component in Step 5 by the sum of the products of all the components.

Step 7: In the first iteration $k = 1$, the mixing proportions π_j for each j^{th} component are estimated by the mean of the initial values of z_{ij} for the corresponding j^{th} component obtained in Step 6.

Step 8: In the first iteration k =1, the mean direction $\mu_j, v_j, \kappa_1, \kappa_2$and λ are estimated by maximising the log-likelihood function given by (14).The "optim" function in R software is used for its maximisation by supplying the initial values of the parameters obtained in Steps 1 and 3 and the normalising constants obtained in Step 2 for each of the p mixture components.

Step 9: The Steps 2 to 8 are repeated until the difference between the values of the log-likelihood function of two consecutive iterations changes by an arbitrary small amount.

The E-step of the EM algorithm on the $(k+1)^{th}$ iteration requires

$$E_{\psi(k)}(z_{ij}|\theta) = P_{\psi(k)}(z_{ij} = 1|\theta) = z_{ij}^{(k)}$$
$$= \frac{\pi_j^{(k)} f_j(\theta i;\eta_j^{(k)})}{f(\theta_i;\psi^{(k)})}; j = 1,...,p,\ i = 1,...,n$$

The M-step on the $(k+1)^{th}$ iteration requires replacing each z_{ij} by $z^{(k)}$ in (*) to give $\pi_j^{(k+1)} = \sum_{i=1}^{n} z_{ij}^{(k)}/n$, $j = 1,...,p$.

The M-step also requires the computation of $\mu_j^{(k)}$, $(j = 1,...,p)$, $\kappa_1^{(k)}$ and $\kappa_{i_j}^{(k)}$ which involves replacing z_{ij} by $z_{ij}^{(k)}$ in the likelihood equations for the estimators, where the parameters with upper suffix k denote the estimators of the parameters at the k^{th} iteration.

The E and M-steps are alternated repeatedly, as in the case of Model 1, till the convergence of the sequence of the log-likelihood values.

3.2 Second stage clustering

Suppose q clusters are obtained in the first stage. The next step is to cluster the observations of each of those q clusters using the mixture model based on the conditional distribution of the other circular variable ϕ given the previous variable θ.

3.2.1 Model 1

Consider the conditional distribution of ϕ_i given θ_i as a von-Mises distribution with mean direction $v(\theta)$ and concentration parameter ρ (θ),where

$$E(\phi|\theta) = v(\theta) = v + 2\arctan(\beta\theta)\ (\mathrm{mod} 2\pi)$$

We assume for simplicity $\rho(\theta) = \rho$ to be constant independent of θ.

The second stage clustering is based on the mixture model

$$f(\phi|\theta;\psi) = \sum_{j=1}^{p} \lambda_j f_j(\phi_i|\theta_i;\eta_j) \qquad (15)$$

where p denotes the number of mixture components, $\eta = (v_1,...,v_p,\beta_1,...,\beta_p,\rho)$, $\psi = (\eta',\lambda_1,...,\lambda_{p-1})$ $\left(\text{since } \lambda_p = 1-\sum_{j=1}^{p-1} \lambda_j\right)$ denotes the parameter vectors of the mixture model, $\eta_j = (v_j,\beta_j,\rho)'$, $(j = 1,...,p)$ denotes the parameter vector of the j^{th} component of the mixture model and

$$f_j(\phi_i|\theta_i;\eta_j) = \frac{1}{2\pi I_0(\rho)} \exp[\rho\cos\{\phi_i - v_j - 2\arctan(\beta_j\theta_i)\}] \qquad (16)$$

Note that here we substitute the values of θ_i of the q clusters obtained in thefirst stage.

3.2.2 Estimation of parameters in Model 1

We use EM algorithm to estimate the parameters of the mixture model. Let y_{ij} be an indicator variable

$$y_{ij} = \begin{cases} 1 & \text{if}(\phi_i|\theta_i) \text{ comes from the } j^{th} \text{ component of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

Using the complete data vector $(y_{ij},\phi_i)'$,the complete data log-likelihood function for $\psi$is given by

$$L_c(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij} \ln f_j(\phi|\theta;\eta_j)$$

$$= -n \ln 2\pi - n \ln I_0(\rho) + \sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij} \left[\rho\cos\{\phi_i - v_j - 2\arctan(\beta_j\theta_i)\}\right] \quad (17)$$

Note that here n denotes the number of observations of each of the q clusters obtained in the first stage.

If the y_{ij}'s are observable, then the MLE of λ_j is simply given by $\hat{\lambda}_j = \sum_{i=1}^{n} y_{ij}/n$.

The likelihood equations for estimating the parameters v_j, (j = 1,. .. ,p), ρ and β_j, (j = 1,...,p) respectively are given as

$$\sum_{i=1}^{n} \hat{y}_{ij}\left[\hat{\rho}\sin\{\phi_i - \hat{v}_j - 2\arctan(\beta_j\theta_i)\}\right] = 0 \quad (18)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{p} \hat{y}_{ij} \cos\{\phi_i - \hat{v}_j - 2\arctan(\beta_j\theta_i)\} - \frac{nI_1(\hat{\rho})}{I_0(\hat{\rho})} = 0 \quad (19)$$

$$\sum_{i=1}^{n} \phi_i\hat{y}_{ij}\left[\hat{\rho}\sin\{\phi_i - \hat{v}_j - 2\arctan(\beta_j\theta_i)\}\right] = 0 \quad (20)$$

Hence solving the likelihood equations(18),(19) and(20),the maximum likelihood estimators of the parameters are given by

$$\hat{v}_j = \arctan^*(\overline{\theta}_2) \qquad j = 1,...,p$$

$$\text{where } \overline{\theta}_2 = \frac{\sum_{i=1}^{n} \hat{y}_{ij} \sin\{\phi_i - 2\arctan(\hat{\beta}_j\theta_i)\}}{\sum_{i=1}^{n} \hat{y}_{ij} \cos\{\phi_i - 2\arctan(\hat{\beta}_j\theta_i)\}} \quad (21)$$

$\hat{\rho} = A^{-1}(\overline{R}_2)$ where

$$\overline{R}_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} \hat{y}_{ij}\cos\{\phi_i - \hat{v}_j - 2\arctan(\hat{\beta}_j\theta_i)\}}{n} \quad (22)$$

The steps of the EM algorithm are presented as follows:

Step 1: A random sample of size of the number of components p generated from uniform distribution over the range of the given data on the circular variable is taken as the initial estimate of a part of the component mean direction v and a random sample of size p generated from a standard normal distribution is taken as the initial estimate of β as another part of the component mean direction. The initial estimate of the concentration parameter ρ is generated from uniform distributions ranging from 0 to the value of the sample concentration of the circular variable. The mixing proportions π are initialized as a random sample of size p from a Dirichlet distribution with shape parameter 1.

Step 2: The conditional distribution used in Stage 2 that is the von-Mises density given by (16) is computed for the observed circular data $(\phi_i|\theta_i)$ (for each cluster obtained in Stage 1) using the initial values of the parameters for each of the p mixture components obtained in Step 1.

Step 3: The product of the mixing proportions π and the densities obtained in Step 2 are calculated for each component.

Step 4: The initial values of the indicator variable y_{ij} are obtained by dividing the product obtained for each j^{th} component in Step 3 by the sum of the products of all the components.

Step 5: In the first iteration $k = 1$, the mixing proportions π_j for each j^{th} component are estimated by the mean of the initial values of y_{ij} fort he corresponding j^{th} component obtained in Step 4.

Step 6: In the first iteration $k = 1$, the mean direction v_j and concentration parameter ρ are estimated by the solutions given by (21) and (22) substituting the initial values of y_{ij} obtained in Step 4.

Step 7: The estimate of the parameter β is updated by maximising the log-likelihood function using the "optim" function in R software by putting in the estimates of the other parameters for each iteration.

Step 8: The log-likelihood equation given by (17) is calculated substituting the initial values of y_{ij}, the observed circular data $(\phi_i|\theta_i)$ and the estimates of the parameters obtained in Steps 6 and 7.

Step 9: The Steps 2 to 8 are repeated until the difference between the values of the log likelihood function of two consecutive iterations changes by an arbitrary small amount.

3.2.3 Model 2

Here the conditional distribution of ϕ given θ is circular normal with mean direction $m = \arctan\frac{\lambda\sin(\theta-\mu)}{\kappa_2}$ and concentration parameter $\kappa = \sqrt{\kappa_2^2 + \lambda^2\sin^2(\theta-\mu)}$ i.e the probability density

function is

$$g(\phi|\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\phi - m)\}$$

The second stage clustering is based on the mixture model

$$f(\phi|\theta;\psi) = \sum_{j=1}^{p} \alpha_j f_j(\phi_i|\theta_i;\eta_j) \quad (23)$$

where p denotes the number of mixture components, $\eta = (m_1,...,m_p,\kappa_1,...,\kappa_p)$, $\psi = (\eta',\alpha_1,...,\alpha_{p-1})$ $\left(\text{since } \alpha_p = 1-\sum_{j=1}^{p-1} \alpha_j\right)$ denote the parameter vectors of the mixture model, $\eta_j=(m_j,\kappa_j)'$, (j= 1,...,p) denote the parameter vector of the j^{th} component of the mixture model and

$$f_j(\phi_i|\theta_i;\eta_j) = \frac{1}{2\pi I_0(\kappa_j)} \exp\{\kappa_j \cos(\phi_i - m_j)\} \quad (24)$$

Note that here we substitute the values of θ_i and the estimators of μ_j,κ_1,λ of the q clusters obtained in the first stage.

3.2.4 Estimation of parameters in Model 2

We use EM algorithm to estimate the parameters of the mixture model. Let y_{ij} be an indicator variable

$$y_{ij} = \begin{cases} 1 & \text{if}(\phi_i|\theta_i) \text{ comes from the } j^{th} \text{ component of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

Using the complete data vector $(y_{ij},\phi_i)'$, the complete data log-likelihood function for ψ is given by

$$L_c(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij} \ln f_j(\phi|\theta;\eta_j)$$

$$= -n \ln 2\pi - \sum_{j=1}^{p} \ln I_0(\kappa_j) + \sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij} \left[\kappa_j \cos\{\phi_i - m_j\}\right] \quad (25)$$

If the y_{ij}' s are observable, then the MLE of α_j is simply given by $\hat{\alpha}_j = \sum_{i=1}^{n} y_{ij}/n$.

The likelihood equations for estimating the parameters (m_j, κ_j, (j = 1,...,p)) respectively are given as

$$\sum_{i=1}^{n} \hat{y}_{ij} \left[\hat{\kappa}_j \sin\{\phi_i - \hat{m}_j\}\right] = 0 \quad (26)$$

$$\sum_{i=1}^{n} \hat{y}_{ij} \cos\{\phi_i - \hat{m}_j\} - \frac{nI_1(\hat{\kappa}_j)}{2\pi I_0(\hat{\kappa}_j)} = 0, \left(I_1(\kappa_j) = \frac{d}{d\kappa_j} I_0(\kappa_j)\right) \quad (27)$$

Hence solving the likelihood equations (26) and (27), the maximum likelihood estimators of the parameters are given by

$$\hat{m}_j = \arctan^*(\bar{\phi}_1) \; j = 1,...,p \text{ where } \bar{\phi}_1 = \frac{\sum_{i=1}^{n} \hat{y}_{ij} \sin\phi_i}{\sum_{i=1}^{n} \hat{y}_{ij} \cos\phi_i} \quad (28)$$

and arctan* defines the quadrant specific angle as given in [3].

$$\hat{\kappa}_j = A^{-1}(\bar{R}) \quad \text{where } \bar{R} = \frac{\sum_{i=1}^{n} \hat{y}_{ij} \cos\{\phi_i - \hat{m}_j\}}{n} \text{ and}$$

$$A(\kappa_j) = \frac{I_1(\kappa_j)}{I_0(\kappa_j)} \quad (29)$$

The steps of the EM algorithm are presented as follows:

Step 1: The initial values of the mean direction m and the concentration parameter κ of the conditional distribution used in Stage 2 for each component are obtained from their relations with the final estimates of the parameters μ, κ_2 and λ obtained for each cluster in Stage 1. The mixing proportions π are initialized as a random sample of size p from a Dirichlet distribution with shape parameter 1.

Step 2: The conditional distribution used in Stage 2 that is the von-Mises density given by (24) is computed for the observed circular data $(\phi_i|\theta_i)$ (for each cluster obtained in Stage 1) using the initial values of the parameters for each of the p mixture components obtained in Step 1.

Step 3: The product of the mixing proportions π and the densities obtained in Step 2 are calculated for each component.

Step 4: The initial values of the indicator variable y_{ij} are obtained by dividing the product obtained for each j^{th} component in Step 3 by the sum of the products of all the components.

Step 5: In the first iteration $k=1$, the mixing proportions π_j for each j^{th} component are estimated by the mean of the initial values of y_{ij} for the corresponding j^{th} component obtained in Step 4.

Step 6: In the first iteration $k=1$, the mean direction m_j and concentration parameter κ_j are estimated by the solutions given by (28) and (29) substituting the initial values of y_{ij} obtained in Step 4.

Step 7: The log-likelihood equation given by (25) is calculated substituting the initial values of y_{ij}, the observed circular data $(\phi_i|\theta_i)$ and the estimates of the parameters obtained in Step 6.

Step 8: The Steps 2 to 7 are repeated until the difference between the values of the log-likelihood function of two consecutive iterations changes by an arbitrary small amount.

4. CLUSTER ALLOCATION IN HIERARCHICAL CLUSTER-ING

4.1 First stage clustering

Suppose there are n observations on the circular random variable θ_i, $i=1,...,n$. Our objective is to allocate these n observations into p homogeneous groups or clusters. We assume that each component of the mixture model to be fitted characterizes the corresponding cluster. In the E-step of the EM algorithm, a $(n\times p)$ matrix is created whose i^{th} row contains estimates of the conditional (on the current parameter estimates) probabilities that the i^{th} observation θ_i belongs to the j^{th} mixture component, i.e. cluster j, $(j=1,...,p)$. On convergence, the i^{th} observation is assigned to the cluster j for which the conditional probability of membership given by

$$P_{i|j}=\frac{\hat{\pi}_j f_j(\theta_i;\hat{\eta}_j)}{\sum_{j=1}^{p}\hat{\pi}_j f_j(\theta_i;\hat{\eta}_j)}$$

is the largest.

4.2 Second stage clustering

Similar to the first stage, here also we need to allocate the n observations on the circular random variable ϕ_i, $i=1,...,n$ into p clusters. We assume that each component of the mixture model to be fitted characterizes the corresponding cluster. In the E-step of the EM algorithm, a $(n\times p)$ matrix is created whose i^{th} row contains estimates of the conditional (on the current parameter estimates) probabilities that the i^{th} observation ϕ_i belongs to the j^{th} mixture component, i.e. cluster j, $(j=1,...,p)$.On convergence, the i^{th} observation is assigned to the cluster j for which the conditional probability of membership given by

$$P_{i|j}=\frac{\hat{\lambda}_j f_j(\phi_i|\theta_i;\hat{\eta}_j)}{\sum_{j=1}^{p}\hat{\lambda}_j f_j(\phi_i|\theta_i;\hat{\eta}_j)}$$

is the largest.

5. CHOICE OF OPTIMUM NUMBER OF CLUSTERS

An important question that arises is how to decide on a reasonable value for the number of clusters. One solution can be obtained by using Bayesian information criterion (BIC) which is given by

$BIC=-2\ln L_{max}+Penalty$

$=-2\ln L_{max}+\ln(n)\times$ number of parameters estimated

$=-2\ln L_{max}+(p+1)\ln(n)$ in case of Model 1 (Stage 1)

$=-2\ln L_{max}+(2p+3)\ln(n)$ in case of Model 2 (Stage 1)

$=-2\ln L_{max}+(2p+1)\ln(n)$ in case of Model 1 (Stage 2)

$=-2\ln L_{max}+2p\ln(n)$ in case of Model 2 (Stage 2)

In order to decide on the number of clusters, we choose that value of p for which the BIC is minimized ([8] and[5]).Another solution can be obtained by minimizing Akaike information criterion(AIC) given by

$AIC=-2\ln L_{max}+2\times$ number of parameters estimated

$=-2\ln L_{max}+2.(p+1)$ in case of Model 1 (Stage 1)

$=-2\ln L_{max}+2.(2p+3)$ in case of Model 2(Stage 1)

$=-2\ln L_{max}+2.(2p+1)$ in case of Model 1 (Stage 2)

$=-2\ln L_{max}+2.2p$ in case of Model 2 (Stage 2)

6. EXAMPLE

[7] present paired observations on the peak expression times or phases (from a micro array

experiment) for 38 circadian related genes common to both heart and liver tissues *in vivo*. This is a popular dataset and we illustrate our methodology through this example. Identification of clusters for these genes can be of interest since the effect of drugs for heart-liver problems may be similar for the genes within each cluster. Here we restrict the number of clusters to at most 6, to avoid sparse clusters due to the size of the data.

6.1 Application of Model 1

The values of BIC and AIC calculated for different values of p for the two stages of clustering in are shown in the following tables 1 and 4 respectively.

Table 1. The values of BIC and AIC for Stage 1

No. of Clusters	BIC	AIC
2	86.48572	81.57296
3	72.60717	66.05682
4	93.76087	85.57294
5	79.88234	70.05683
6	83.51993	72.05683

It can be observed from Table 1 that the BIC and the AIC criterion are minimum for p=3 clusters as shown in Fig.1.

Then following the clustering technique as described before, the 38 observations are assigned to the 3 clusters as displayed in Table 2.

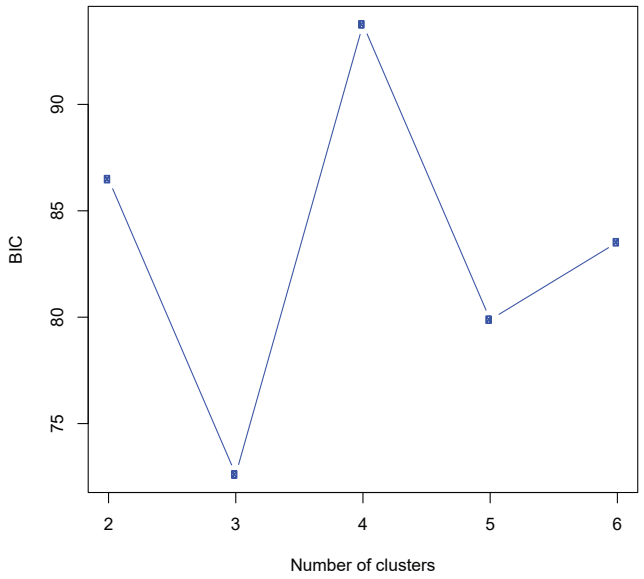

Fig. 1. Plot of BIC for first stage clustering for phase in heart under Model 1

Table 2. Cluster Allocation (Stage 1)

Cluster Number	No. of Observations	Observation Number
1	4	10-13
2	14	14-27
3	20	1-9,28-38

Table 3. Results from EM Algorithm for the optimum number of clusters (Stage 1)

No. of Iterations	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\pi}$	BIC
111	$\begin{pmatrix} 2.38590964 \\ -2.41876525 \\ -0.04568916 \end{pmatrix}$	4.281714	$\begin{pmatrix} 0.08745852 \\ 0.38069268 \\ 0.53184880 \end{pmatrix}$	72.60717

It can be observed from Table 4 that the BIC and the AIC criterion are minimum for $p = 2$ clusters for the cluster 1 obtained in Stage 1. For cluster 2 obtained in Stage 1, the BIC and the AIC criterion are minimum for $p = 2$ clusters but observations are allocated in only one cluster due to one of the small mixing proportions. The next minimum are for $p = 3$ clusters but that also suffers from the same problem as for $p = 2$ clusters. The next minimum are for $p = 5$ clusters but again observations are allocated only in 3 clusters. Hence, the optimum number of clusters here is taken as 3 as the differences between the BIC and AIC values corresponding to $p = 3$ and $p = 5$ are quite small (as can be seen from Fig. 2) although the observations are allocated in only 2 clusters due to one of the small mixing proportions. For cluster 3 obtained in first stage, the minimum AIC and BIC values are obtained for $p = 6$ clusters.

Table 4. The values of BIC and AIC for Stage 2

Cluster no. in Stage 1	Cluster 1		Cluster 2		Cluster 3	
No. of Clusters	BIC	AIC	BIC	AIC	BIC	AIC
2	1.438447	4.506975	40.55766	37.36237	71.54768	66.56902
3	18.29015	22.58609	98.35981	93.88641	44.12541	37.15528
4	88.93583	94.45918	208.2556	202.504	56.18758	47.22599
5	–	–	103.2881	96.25847	83.98067	73.02761
6	–	–	528.9134	520.6057	33.78317	20.83865

Thus the allocation of the observations to the sub-clusters in Stage 2 are shown in Table 5 below.

Table 5. Cluster Allocation (Stage 2)

Cluster Number In Stage 1	Sub-Cluster Number In Stage 2	No. of Observations	Observation Number
1	1	1	13
	2	3	10-12
2	1	10	14-18,20-23,27
	2	4	19,24-26
3	1	4	1,4,6,30
	2	1	35
	3 4	1 5	3 2,9,32-33,38
	5	5	7-8,28-29,34
	6	4	5,31,36-37

Table 6. Results from EM Algorithm for the optimum number of clusters (Stage 2)

Cluster Number	No. of Iterations	$\hat{\nu}$	$\hat{\rho}$	$\hat{\beta}$	BIC
1	21	$\begin{pmatrix}-0.7024817\\1.5386203\end{pmatrix}$	66.98927	$\begin{pmatrix}0.6342904\\0.0119862\end{pmatrix}$	1.438447
2	84	$\begin{pmatrix}-0.5603809\\-0.2043977\end{pmatrix}$	8.785871	$\begin{pmatrix}500669.099318\\-0.742157\end{pmatrix}$	40.55766
3	194	$\begin{pmatrix}0.5541931\\-0.9237106\\-2.4443983\\1.2368153\\0.7772537\\-0.2896228\end{pmatrix}$	22.59374	$\begin{pmatrix}-0.34930741\\-0.08928988\\1.33086609\\-0.16403511\\1.42497165\\-0.48611296\end{pmatrix}$	33.78317

6.2 Application of Model 2

The values of BIC and AIC are calculated for different values of p for the two stages in the following tables 7 and 10 respectively.

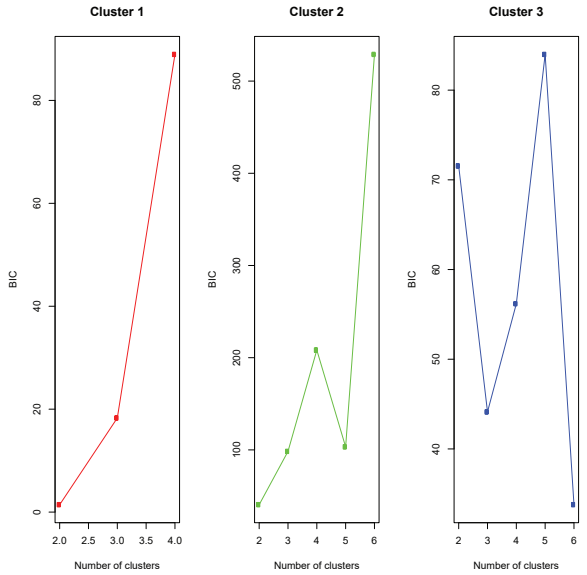

Fig. 2. Plot of BIC for second stage clustering of clusters 1, 2 and 3 respectively obtained in first stage for phase in liver under Model 1

Table 7. The values of BIC and AIC for Stage 1

No. of Clusters	BIC	AIC
2	149.2601	137.797
3	156.9873	142.249
4	159.7166	141.7032
5	173.0366	151.748

It can be observed from Table 7 that the BIC and the AIC criterion are minimum for p=2 clusters as shown in Fig. 3.

Then following the clustering technique as described before, the 38 observations are assigned to the 3 clusters as displayed in Table 8.

Table 8. Cluster Allocation (Stage1)

Cluster Number	No. of Observations	Observation Number
1	11	8-9,17-25
2	27	1-7,10-16,26-38

Table 9. Results from EM Algorithm for the optimum number of clusters (Stage 1)

No. of Iterations	$\hat{\mu}$	$\hat{\kappa}_1$	$\hat{\pi}$	BIC
2	$\begin{pmatrix}-1.302621\\1.663844\end{pmatrix}$	0	$\begin{pmatrix}0.4682564\\0.5317436\end{pmatrix}$	149.2601

Table 10. The values of BIC and AIC for Stage 2

Cluster no. in Stage 1	Cluster 1		Cluster 2	
No. of Clusters	BIC	AIC	BIC	AIC
2	40.96793	39.37635	106.7313	101.5479
3	45.7637	43.37633	113.3812	105.6061
4	50.55951	47.37635	119.9716	109.6049
5	55.3553	51.37634	126.5895	113.6311
6	60.15107	55.37633	133.1594	117.6093

It can be observed from Table 10 that the AIC and BIC values are minimum for p = 2 clusters for both the clusters 1 and 2 obtained in Stage 1. But the observations are allocated in only one cluster in Stage 2 for both the clusters in Stage 1.

Thus the allocation of the observations to the sub-clusters in Stage 2 are shown in Table 11 below.

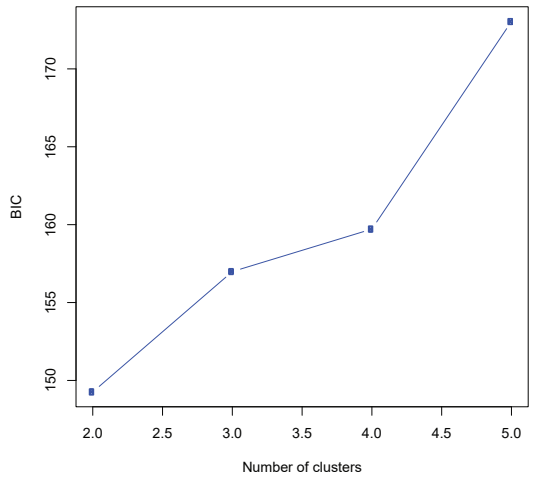

Fig. 3. Plot of BIC for first stage clustering for phase in heart under Model 2

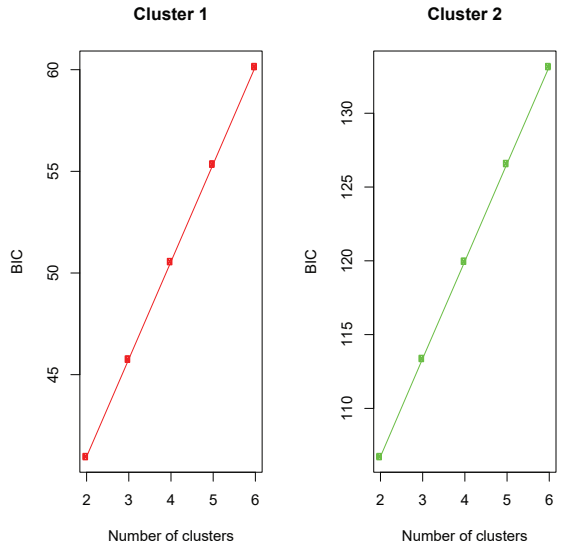

Fig. 4. Plot of BIC for second stage clustering of clusters 1 and 2 respectively obtained in first stage for phase in liver under Model 2

Table 11. Cluster Allocation (Stage 2)

Cluster Number In Stage 1	Sub-Cluster Number In Stage 2	No. of Observations	Observation Number
1	1	11	8-9,17-25
2	1	27	1-7,10-16,26-38

Table 12. Results from EM Algorithm for the optimum number of clusters (Stage 2)

Cluster Number	No. of Iterations	$\hat{m}$	$\hat{\kappa}$	$\hat{\pi}$	BIC
1	59	$\begin{pmatrix}2.62554\\-3.00402\end{pmatrix}$	$\begin{pmatrix}1.924783e{-}05\\1.537110e{+}00\end{pmatrix}$	$\begin{pmatrix}2.841376e{-}05\\9.999716e{-}01\end{pmatrix}$	40.96793
2	199	$\begin{pmatrix}-2.8752806\\-0.2591533\end{pmatrix}$	$\begin{pmatrix}0.02172818\\0.62619506\end{pmatrix}$	$\begin{pmatrix}0.1454958\\0.8545042\end{pmatrix}$	106.7313

7. COMPARISON OF PERFORMANCE OF THE CLUSTERING UNDER MODELS 1 AND 2

The performance of the clustering is assessed through the computation of the AIC and BIC values using the maximum values of the complete log-likelihood function of all the parameters of both the stages of clustering. The maximum complete log-likelihood function is based on the independent sums of the products of the marginal distribution of the first variable used in the first stage and the conditional distribution of the second variable given the first variable used in the second stage of clustering. The observations allocated to the optimum number of clusters and the estimates of the parameters obtained in the optimum number of clusters are substituted to obtain the maximum log-likelihood function. Thus from equations (4) and (17), the estimated complete log-likelihood function under Model 1 substituting the EM algorithm estimates is given by

$$Lmax = \sum_{j=1}^{p_1}\left[\left[-n_{1_j}\ln 2\pi - n_{1_j}\ln I_0(\hat{\kappa}) + \sum_{i=1}^{n_{1_j}} z_{ij}\left\{\hat{\kappa}\cos\{\theta_i - \hat{\mu}_j\}\right\}\right]\sum_{k=1}^{p_2}\left[-n_{2_k}\ln 2\pi - n_{2_k}\ln I_0(\hat{\rho}) + \sum_{l=1}^{n_{2_k}} y_{kl}\left\{\hat{\rho}\cos\{\phi_k - \hat{\nu}_l - 2\arctan(\hat{\beta}_l\theta_k)\}\right\}\right]\right] \quad (30)$$

Also from equations (14) and (25), the estimated complete log-likelihood function under Model 2 substituting the EM algorithm estimates is given by

$$Lmax = \sum_{j=1}^{p_1}\left[\left[n_{1_j}\ln 2\pi + \sum_{i=1}^{n_{1_j}} z_{ij}\left\{\ln C_j + \ln I_0(\hat{\kappa}_{i_j}) + \kappa_1\cos\{\theta_i - \hat{\mu}_j\}\right\}\sum_{k=1}^{p_2}\left[-n_{2_k}\ln 2\pi - \sum_{l=1}^{n_{2_k}} y_{kl}\left\{\hat{\kappa}_l\cos\{\phi_k - \hat{m}_l\}\right\}\right]\right] \quad (31)$$

where p_1 and p_2 denote the optimum number of clusters in Stages 1 and 2 respectively, n_{1_j} and n_{2_k} denote the number of observations in the j^{th} and k^{th} cluster in Stages 1 and 2 respectively.

For the above data on genes of heart and liver tissues, substituting the values of p_1=3 and p_2=2,2,6 and the EM algorithm estimates of the parameters from Tables 3 and 6 for Model 1, its values of AIC and BIC of Model 1 are obtained below. Similarly, substituting the values of p_1=2 and p_2=1,1 and the EM algorithm

estimates of the parameters from Tables 9 and 12 for Model 2, its values of AIC and BIC are also obtained in the following table.

Table 13. The values of BIC and AIC for the complete model

Model 1		Model 2	
BIC	**AIC**	**BIC**	**AIC**
-9525.029	-9542.625	-4784.684	-4794.984

It can be observed from Table 13 that Model 1 yields smaller values of AIC and BIC than Model 2. Hence the performance of the clustering under Model 1 can be preferred to that under Model 2.

8. CONCLUDING REMARKS

The newly introduced method of hierarchical clustering has been applied to the toroidal data on the peak expression times of the heart and liver tissues under two mixture models viz. Model 1 and Model 2 as defined above. The cluster allocations are obtained for stages 1 and 2 based on the marginal and the conditional distributions respectively under both models 1 and 2. The overall performance of the clustering is assessed by combining the sums of the mixture distributions of the independent components of the optimum number of clusters of both the stages under both the models. It is found that Model 1 performs better than Model 2 in the clustering technique for the data based on the AIC and BIC values of the combined sum of the mixture distributions.

ACKNOWLEDGEMENTS

The first author would like to thank the Editors for inviting him to submit a paper in this volume honouring Prof. Prem Narain, with whom he had the opportunity to interact and also admire his profound research contributions.

REFERENCES

Dingxi, Q. and Tamhane, A. (2007). A Comparative Study of K-means Algorithm and the Normal Mixture Model for Clustering: Univariate Case, *Journal of Statistical Planning and Inference*. **137**, 3722-3740.

Fisher, N.I., and Lee, A.J. (1992). Regression Models for an Angular Response, *Biometrics*, **48**(3), 665-677.

Jammalamadaka, S.R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific Publishers, New Jersey.

Johnson, R.A., and Wehrly, T.E. (1978). Some angular-linear distributions and related regression models, *Journal of the American Statistical Association*, **73**, 602-606.

Johnson, R.A., and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, New Jersey.

Kim, S. and SenGupta, A. (2017). Multivariate Multiple Circular Regression. *Journal of Statistical Computation and Simulation.* **87**, 1277-1291.

Liu, D., Peddada, S.D., Li, L. and Weinberg, C.R. (2006). Phase analysis of circadian-related genes in two tissues. *Bioinformatics BMC*, **7**(87), https://doi.org/10.1186/1471-2105-7-87.

McLachlan, G.J., and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, NewYork.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, Second Edition. John Wiley & Sons, NewYork.

SenGupta, A. (2004). On the Construction of Probability Distributions for Directional Data. *Bul. Cal. Math. Soc.*, **96**(2), 139-154.

SenGupta, A., Arnold, B. and CandKim, S. (2013). Inverse Circular-Circular Regression. *Journal of Multivariate Analysis.* **119**, 200-208.

SenGupta, A., and Roy, M. (2021). Clustering on the Torus. *Journal of Statistical Theory and Practice,* **58**, 1-18.

SenGupta, A., and Ugwuowo, F. (2011). A Classification Method for Directional Data with Application to the Human Skull. *Communication in Statistics, Theory and Methods*, **40**, 457-466,.

Singh H., Hnizdo V., and Demchuk E.(2002). Probabilistic model for two dependent circular variables. *Biometrika*, **89**(3), 719-723.

Wu, J.C.F. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics*, **11**(1), 95-103.