Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 51 (4) (2022), 1160 – 1173
DOI : 10.15672/hujms.897144

Research Article

A flexible Bayesian mixture approach for multi-modal circular data

Muhammet Burak Kılıç*[1], Zeynep Kalaylioglu[2], Ashis SenGupta[2,3,4]

[1] *Department of Business Administration, Quantitative Methods Unit, Burdur Mehmet Akif Ersoy University, Burdur, Turkey*
[2] *Department of Statistics, Middle East Technical University, Ankara, Turkey*
[3] *Indian Statistical Institute, Kolkata, India*
[4] *Department of Biostatistics and Epidemiology, Augusta University, Augusta, Georgia, USA*

Abstract

In this article, we consider multi-modal circular data and nonparametric inference. We introduce a doubly flexible method based on Dirichlet process circular mixtures in which parameter assumptions are relaxed. We assess and discuss in simulation studies the efficiency of the proposed extension relative to the standard finite mixture applications in the analysis of multi-modal circular data. The real data application shows that this relaxed approach is promising for making important contributions to our understanding of many real-life phenomena particularly in environmental sciences such as animal orientations.

Mathematics Subject Classification (2020). 62G05, 62P12, 97K80

Keywords. directional data, Dirichlet process prior, mixture models, stick breaking construction, animal orientation

1. Introduction

In many sciences including environmental, biological, pharmaceutical, geophysical and astrophysical, the data are directional, e.g. orientations of animals such as birds and turtles, wind directions, orientation of geological deposits and arrival directions of the ultra high energy cosmic rays. Focus of the current article is univariate circular data. These are directional data observed on the unit circle. Most circular data are multi-modal (i.e. number of modes ≥ 2), [24, 25]. For instance, orientation behaviour of migratory nocturnal songbirds, Sylvia communis in particular, is multi-modal. They are shown to prefer seven different migration routes ranging from 39.8^o to 326.6^o clockwise from North with standard deviations ranging from 5^o to 14.4^o [23]. Recent literature is abundant with examples on multi-modal avian migration routes preferred by some other species as well [2, 10]. Other examples include the turtle orientations given in Section 4.

Fisher [6] (p. 96) lists the main questions regarding multi-modal circular data: what is the number of unimodal populations present in the sample; what are the mean directions

---

*Corresponding Author.

Email addresses: mburak@mehmetakif.edu.tr (M.B. Kılıç), kzeynep@metu.edu.tr (Z. Kalaylioglu), amsseng@gmail.com (A. SenGupta)

Received: 02.04.2021; Accepted: 31.03.2022

of these groups; what are the relative proportions of the groups? Appropriate statistical analysis of multi-modal circular data should address these questions and distinguish distributions (clusters/groups) with different mean and concentration (or dispersion) properties present in the data. For bimodal circular data, where number of modes is equal to 2, generalized von Mises [8, 31] and asymmetric generalized von Mises [17] distributions can be used. Various types of multi-modal circular data can be analyzed using mixture of unimodal symmetric circular distributions, particularly mixture of von Mises (vM) and mixture of wrapped Cauchy (wC) distributions [11,16,19]. In finite mixtures, multi-modal circular density is a linear combination of finite number of unimodal circular densities. Letting $\theta \in [0, 2\pi)$ or $\theta \in [-\pi, \pi)$ be a circular random variable, probability density function of circular mixture distribution is given as follows

$$f(\theta; p_1, ..., p_C, \varphi_1, \varphi_2, ..., \varphi_C) = \sum_{k=1}^{C} p_k f_k(\theta; \varphi_k), \tag{1.1}$$

where C is the number of distributions in the mixture, $0 \leq p_k \leq 1$ for $k = 1, ..., C$ are unknown mixing probabilities subject to $\sum_{k=1}^{C} p_k = 1$, f_k is the probability density function of the population k, and φ_k is the vector of parameters associated with the kth mixing distribution. The p_k is proportion of that particular density in the overall mixture distribution. In the finite mixture approach, number of mixing components is assumed known and C is set fixed. A general account of identifiability of finite circular mixture distributions is given by [12] (Theorem 2.1. on p. 441) where identifiability for vM mixtures in particular is given by [7].

Nonparametric mixture approaches provide a flexible alternative to finite mixtures because number of mixing components need not be known in advance and C is left unspecified. Dirichlet process (DP) mixture models, in which $C \to \infty$, is the most popular nonparametric approach particularly for data on Euclidean space. Its adoption for circular data has been difficult due to the computational demands mostly arising from the normalizing constants of the popular circular distributions that are not usually available in closed forms. These difficulties have been alleviated by introducing assumptions about the parameters of mixing densities. One assumption is to use completely known prior distributions for the concentration parameters [4, 9]. Another assumption is homogeneity of concentration parameters of mixing distributions [21, 22]. These approaches are theoretically and computationally attractive and used in many applications. On the other hand concentration parameters of different populations present in the sample need not be the same [16] and there are recent examples, e.g. [23], where dispersions (concentration) of underlying groups in the data are different. In the current article, we relax these assumptions and provide a doubly flexible DP approach to analyze the most common circular distributions. To the best of our knowledge, this is the first paper to study DP circular mixture models with different mixing circular precision parameters assuming unknown priors. The contribution of our approach yields simultaneously both the number of underlying populations as well as the estimates of their corresponding parameters. These populations define the underlying clusters. This paper also enables that the problem of determining the optimal number of clusters, the critical step of any usual clustering approach and which is often done through an optimality criterion based on, e.g. information criterion like generalized Akaike information criterion or Bayesian information criterion, is no longer needed here. Thus the only problems remaining are that of identifying the members of these clusters and the efficiency of this clustering approach. Following model-based clustering for toroidal data [26,27] (see also [15]), the identification is here done by the DP circular mixture models where each observation is allocated by its membership measure and the efficiency of our method is determined by the misclassification rate. These are presented in details in the relevant section of our paper. Rest of the article is organized as follows. Section 2 gives a brief review of DP mixture approach. Section 3 gives our doubly flexible DP mixture model for the analysis of multi-modal circular data using well known circular distributions, an efficient Bayesian method for inference is also provided in this section. In Section 4, efficiency of the proposed model is evaluated using extensive numerical studies including a set of hypothetical data, a Monte Carlo study and a real data. Section 5 gives concluding remarks and ends with possible extensions.

2. A review of DP mixture approach

Let G denote an unknown cumulative distribution function. A suitable prior for G is $G \sim \mathrm{DP}(G_0, \alpha)$ where DP is Dirichlet process prior, G_0 is a suitably chosen baseline distribution over the space Ω and α is the precision parameter that shows whether G is in the close realisation of G_0. Letting $(A_1, ..., A_k)$ be a partition of Ω such that $A_1 U ... U A_k = \Omega$, that representation is equivalent to $(G(A_1), ..., G(A_k)) \sim \mathrm{Dir}(\alpha G_0(A_1), ..., \alpha G_0(A_k))$ where Dir denotes Dirichlet distribution [5]. Mixture of Dirichlet processes, i.e. DP mixture models, is a Bayesian nonparametric method in which prior distribution of parameters are specified using DP [3].

Adoption of DP mixture approach for multi-modal circular response variables is as follows. Suppose $\theta_i, i = 1, 2, ..., n$ are circular random variables of size n from a mixture distribution $F(\varphi)$ with parameter vector φ and G is an unknown joint prior distribution for φ. A basic DP mixture model representation is then

$$\begin{aligned} \theta_i | \varphi &\sim F(\varphi) \text{ for } i = 1, ..., n \\ \varphi | G &\sim G \\ G &\sim \mathrm{DP}(G_0, \alpha). \end{aligned} \tag{2.1}$$

Equivalent model can be obtained by taking the limit of the following C component model as C goes to infinity, [20]

$$\begin{aligned} \theta_i | K_i, \boldsymbol{\varphi} &\sim F(\varphi_{K_i}) \\ K_i | \boldsymbol{p} &\sim \mathrm{Discrete}(p_1, ..., p_C) \\ \varphi_k &\sim G_0, \quad k = 1, ..., C \\ \boldsymbol{p} &\sim \mathrm{Dir}(\alpha/C, ..., \alpha/C), \end{aligned} \tag{2.2}$$

where K_i denotes the latent class the observation θ_i belongs to, φ_k is the vector of parameters characterizing the distribution of class (component) k and $\boldsymbol{p} = (p_1, ..., p_C)$ are unknown mixing proportions of classes that are assigned Dirichlet prior with parameters each set at α/C.

DP mixture in Equation (2.2) can be constructed using the stick breaking priors introduced by [28]. Accordingly, for $q_j \sim \mathrm{Beta}(1, \alpha)$, letting $p_1 = q_1$, and $p_k = q_k \prod_{j=1}^{k-1}(1 - q_j)$ for $k = 2, ..., \infty$ define $G = \sum_{k=1}^{\infty} p_k \delta_{\varphi_k}$ where $\varphi_k \sim G_0$ and δ_{φ_k} is a probability density function degenerate at φ_k. It is shown that an accurate approximation to G can be achieved by truncating the infinite sum at a finite integer C [14]. A theoretical method for truncation is also provided therein. In practice, C is initialized at a reasonably large integer and accuracy of the approximation is determined using the posterior distribution of the number of clusters (see, Figure 2). Probability density function centering around numbers smaller than the initial value indicates sufficient approximation is achieved.

3. DP mixture circular models

In this section, a flexible DP mixture model is presented for mixtures of vM and mixtures of wC distributions, the standards for medium and heavy tail circular distributions in most applied problems.

3.1. DP mixture von Mises model

Consider the following model where $\mu_k \epsilon [0, 2\pi)$ or $\mu_k \epsilon [-\pi, \pi)$ and $\kappa_k \epsilon [0, \infty)$ are location and concentration parameters respectively for the kth mixing vM distribution.

$$\begin{aligned} \theta_i | K_i, \boldsymbol{\varphi} &\sim \mathrm{vM}(\mu_{K_i}, \kappa_{K_i}), \quad i = 1, ..., n \\ K_i | \boldsymbol{p} &\sim \mathrm{Discrete}(p_1, ..., p_C) \\ \varphi_k = (\mu_k, \kappa_k) &\sim G_0, \;\; k = 1, ..., C \\ \alpha &\sim \mathrm{Gamma}(v_1, v_2). \end{aligned} \tag{3.1}$$

Here G_0 is an appropriate bivariate distribution such as $G_0 = \mathrm{vM}(\mu_0, \kappa_0) \otimes \mathrm{Gamma}(a_0, b_0)$, $\boldsymbol{\varphi} = (\varphi_{K_1}, ..., \varphi_{K_n})$, and $\boldsymbol{K} = (K_1, ..., K_n) \in (1, ..., C)^n$ under G. The parameters of the Gamma density used throughout the article are shape and rate parameters respectively. Stick breaking priors are assigned for the unknown weights $\boldsymbol{p} = (p_1, ..., p_C)$.

A suitable prior distribution for α is Gamma distribution. Its parameters values are chosen so that they agree with the desired truncation. Larger (smaller) prior variance is associated with truncation at a larger (smaller) number.

3.2. DP mixture wrapped Cauchy model

In a similar fashion, a flexible DP mixture wC model is defined as follows:

$$\begin{aligned} \theta_i | K_i, \boldsymbol{\varphi} &\sim \mathrm{wC}(\mu_{K_i}, \rho_{K_i}), \quad i = 1, ..., n \\ K_i | \boldsymbol{p} &\sim \mathrm{Discrete}(p_1, ..., p_C) \\ \varphi_k = (\mu_k, \rho_k) &\sim G_0, \;\; k = 1, ..., C \\ \alpha &\sim \mathrm{Gamma}(v_1, v_2), \end{aligned} \tag{3.2}$$

where $\mu_k \epsilon [0, 2\pi)$ or $\mu_k \epsilon [-\pi, \pi)$ and $\rho_k \epsilon [0, 1)$ are location and concentration parameters of the kth mixing wC distribution respectively and G_0 is an appropriate bivariate distribution for them e.g. $G_0 = \mathrm{vM}(\mu_0, \kappa_0) \otimes \mathrm{Beta}(a_0, b_0)$ where the hyperparameters are fixed.

3.3. Inference via Gibbs sampling

To obtain direct inference for G, we adopt blocked Gibbs sampling approach described in [14]. To avoid label switching problem, an identifiability constraint, $\kappa_1 \leq \kappa_2 \leq ... \leq \kappa_C$ for DP vM mixtures and $\rho_1 \geq \rho_2 \geq ... \geq \rho_C$ for DP wC mixtures are used to separate the labels subspace. Letting g denote a posterior density, posterior distribution of DP mixture vM model is given below

$$\begin{aligned} g(\boldsymbol{\varphi}, \boldsymbol{p}, \boldsymbol{K}, \alpha | \boldsymbol{\theta}) \propto & \prod_{i=1}^{n} \mathrm{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \mathrm{vM}(\mu_{K_i} | \mu_0, \kappa_0) \times \mathrm{Gamma}(\kappa_{K_i} | a_0, b_0) \\ & \times \prod_{i=1}^{n} \mathrm{Discrete}(K_i | p_C) \times \pi(p_C = p_C(q_{C-1})) \times \mathrm{Gamma}(\alpha | v_1, v_2), \end{aligned} \tag{3.3}$$

where π denotes the prior distribution of $\boldsymbol{p}$ which is obtained by stick breaking algorithm with Beta priors. Marginal posterior distributions are not available in closed forms and thus Gibbs sampling is used for the marginal posterior inference. Full conditional densities for Gibbs sampling are given below

$$g(\boldsymbol{\varphi} | \boldsymbol{K}, \boldsymbol{p}, \alpha, \boldsymbol{\theta}) = g(\boldsymbol{\varphi} | \boldsymbol{K}, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} \mathrm{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \mathrm{vM}(\mu_{K_i} | \mu_0, \kappa_0) \times \mathrm{Gamma}(\kappa_{K_i} | a_0, b_0),$$

$$g(\boldsymbol{K} | \boldsymbol{p}, \boldsymbol{\varphi}, \alpha, \boldsymbol{\theta}) = g(\boldsymbol{K} | \boldsymbol{p}, \boldsymbol{\varphi}, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} \mathrm{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \mathrm{Discrete}(K_i | p_C),$$

$$
\begin{aligned}
g(\boldsymbol{p}|\boldsymbol{K},\boldsymbol{\varphi},\alpha,\boldsymbol{\theta}) &= g(\boldsymbol{p}|\boldsymbol{K},\alpha) \propto \prod_{i=1}^{n} \text{Discrete}(K_i|p_C) \times \pi(p_C = p_C(q_{C-1})) \times \text{Gamma}(\alpha|v_1,v_2), \\
g(\alpha|\boldsymbol{p},\boldsymbol{\varphi},\boldsymbol{K},\boldsymbol{\theta}) &= g(\alpha|\boldsymbol{p}) \propto \pi(p_C = p_C(q_{C-1})) \times \text{Gamma}(\alpha|v_1,v_2).
\end{aligned} \tag{3.4}
$$

This procedure generates samples from the joint posterior distribution $g(\boldsymbol{\varphi},\boldsymbol{p},\boldsymbol{K},\alpha|\boldsymbol{\theta})$. Upon convergence of the chain, $(\boldsymbol{\varphi}^*,\boldsymbol{p}^*)$ produced in each Gibbs cycle are samples of $(\boldsymbol{\varphi},\boldsymbol{p})$ from the joint posterior. These samples generate a random probability measure as shown below:

$$
G^*(.) = \sum_{k=1}^{C} p_k^* \delta_{\varphi_k^*}(.), \tag{3.5}
$$

where G^* can be used directly to estimate the posterior distribution $G|\boldsymbol{\theta}$. The Gibbs algorithm starts from initial values $(\boldsymbol{\varphi}^{(0)},\boldsymbol{p}^{(0)},\boldsymbol{K}^{(0)},\alpha^{(0)})$ and simulates $(\boldsymbol{\varphi}^{(t)},\boldsymbol{p}^{(t)},\boldsymbol{K}^{(t)},\alpha^{(t)})$ from the conditional distributions in Equation (**??**). To initialize the parameters in the Gibbs algorithm, we let a pilot Markov chain Monte Carlo (MCMC) run and used the values in the final iteration as initials for the actual MCMC run. Full conditionals for DP mixture wC model are defined in a similar manner. These are given in the Appendix.

4. Numerical studies

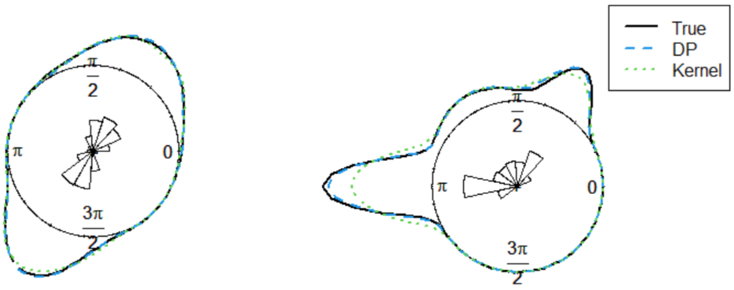
This section provides an evaluation of the performances of the flexible DP circular mixture approaches from various different aspects. Main questions are identifiability of the parameters and properties of the estimators. These questions are investigated using hypothetical as well as real datasets. Circular [1] and R2OpenBUGS [30] packages of R are used for circular data generation and MCMC algorithms respectively. To ensure the MCMC convergence, dynamic traces of Gibbs sampling are monitored and Brooks-Gelman-Rubin's diagnostics are used. To assess the model fit, we used circular kernel density estimation in R. Circular package of R (functions bw.nrd and bw.cv.ml) is used for the bandwidth selection in circular kernel density approximation.

4.1. A simulation study for identifiability

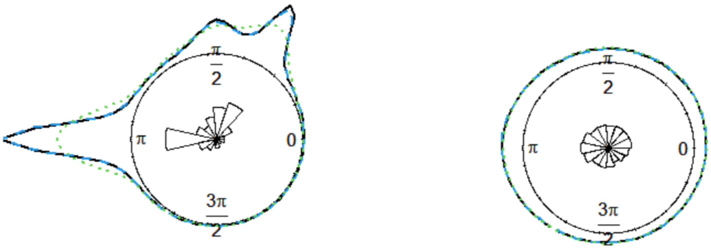
In this section identifiability of the parameters is examined, which may be a potential risk particularly when not only the location parameters but all the parameters of the mixing distributions are component specific and infinite sum of DP mixture is approximated by a rather large C.

We use a representative set of datasets common in practice. Their circular histograms are given in Figure 1. In the figure, the data given in panel (a) show tendency to two main directions (Type 1); data in panel (b) show high concentration around two distinct peaks but the number of peaks is somewhat obscure (Type 2); data in panel (c) show tendency to multiple directions with possibly heavy tails (Type 3); data in panel (d) are almost uniformly distributed (Type 4). Following models are used to generate the multi-modal circular data (in radian for μ) given in these panels respectively, for $i = 1, 2, ..., 1000$:

$$
\begin{aligned}
\theta_i &\sim 0.50 \text{ vM}(\pi/3, 5) + 0.50 \text{ vM}(4\pi/3, 10), \\
\theta_i &\sim 0.25 \text{ vM}(2\pi/3, 5) + 0.25 \text{ vM}(\pi/3, 30) + 0.50 \text{ vM}(\pi, 40), \\
\theta_i &\sim 0.25 \text{ wC}(\pi/3, 0.90) + 0.25 \text{ wC}(\pi/2, 0.80) + 0.50 \text{ wC}(\pi, 0.90), \\
\theta_i &\sim 0.25 \text{ wC}(\pi/3, 0.20) + 0.25 \text{ wC}(\pi/6, 0.30) + 0.50 \text{ wC}(\pi, 0.40).
\end{aligned}
$$



(a) Type 1 (b) Type 2



(c) Type 3 (d) Type 4

Figure 1. Sample circular histograms, DP mixture fitted models, and kernel densities for a representative set of multi-modal circular data.

Baseline distributions for DP priors for the parameters φ_k, $k = 1, ..., C$ are set as $G_0 = \mathrm{vM}(0.01, 7) \otimes \mathrm{Gamma}(1, 1)$ and $G_0 = \mathrm{vM}(0.01, 7) \otimes \mathrm{Beta}(1, 1)$ for analysis with DP mixture vM and DP mixture wC respectively. Prior distribution for α is $\alpha \sim \mathrm{Gamma}(2, 2)$. We set $C = 10$. For corresponding datasets, the results of DP mixture vM and DP mixture wC are presented here. In the MCMC analysis, first 25000 iterations are burnt in and the next 25000 iterations are used for Bayesian calculations. Using the generic notations β, β_0, and D_{obs} for a parameter, its true value, and observed data vector respectively, posterior central location and posterior uncertainty, $F^{-1}_{\beta|D_{obs}}(0.5)$ and $\sqrt{\mathrm{Var}[\beta|D_{obs}]}$, are given in Table 1, the later in parentheses. According to the results, the mixture analysis is, in general, able to identify the underlying clusters and their parameters. As expected, its performance improves if the data present rather distinct clusters (as in the case of Type 2). The analysis well identifies the underlying number of clusters and their parameters even for well mixed data (as in Type 4). Also, Figure 1 shows the true data generation process and the DP mixture fit as well as a kernel estimation for the density. For clustering purposes, the proposed DP circular mixture models are compared to the K-means and spherical K-means [13] clustering algorithms for Type 1, Type 2 and Type 3 examples. Here, the result of Type 4 example is not presented due to the uncertainty between true and estimated clusters. The K-means and the spherical K-means clustering algorithms are implemented in R software. The misclassification rates are obtained by using the confusion matrix and these misclassification rates are presented in Table 2. The misclassification rates of the proposed DP circular mixture models are also close to the results of K-means and spherical K-means clustering and have competitive performances without requiring the optimal cluster number. The posterior distributions of number of clusters based on Type 1, Type 2 and Type 3 examples are presented in Figure 2. As shown, the DP mixture circular methods also well identifies the true number of clusters. Accordingly, not only does the DP circular mixture methods provides a satisfactory fit but also provides class identification.

Table 1. Posterior estimates (posterior uncertainty) for the four types of circular datasets.

	Type 1			Type 2		
k	$\hat{p}_k$	$\hat{\mu}_k$	$\hat{\kappa}_k$	$\hat{p}_k$	$\hat{\mu}_k$	$\hat{\kappa}_k$
1	0.49(0.02)	1.06(0.03)	5.69(0.39)	0.24(0.02)	2.06(0.05)	5.52(1.11)
2	0.50(0.02)	4.20(0.01)	10.48(0.66)	0.26(0.02)	1.07(0.02)	26.51(3.04)
3	0.01(0.01)	0.18(0.40)	17.22(6.05)	0.50(0.02)	3.14(0.01)	36.92(2.61)
4	0.00(0.01)	0.06(0.39)	25.51(8.61)	0.00(0.00)	0.10(0.45)	44.05(6.65)
5	0.00(0.00)	0.03(0.39)	33.48(10.53)	0.00(0.00)	0.04(0.39)	51.96(9.04)
6	0.00(0.00)	0.03(0.38)	41.45(12.14)	0.00(0.00)	0.03(0.38)	60.14(10.91)
7	0.00(0.00)	0.02(0.38)	49.59(13.54)	0.00(0.00)	0.02(0.39)	68.11(12.54)
8	0.00(0.00)	0.00(0.38)	57.40(14.79)	0.00(0.00)	0.01(0.37)	76.08(13.92)
9	0.00(0.00)	0.02(0.38)	65.52(16.00)	0.00(0.00)	0.03(0.40)	85.04(15.18)
10	0.00(0.00)	0.02(0.38)	73.41(17.12)	0.00(0.00)	0.01(0.37)	92.98(16.38)
	Type 3			Type 4		
k	$\hat{p}_k$	$\hat{\mu}_k$	$\hat{\rho}_k$	$\hat{p}_k$	$\hat{\mu}_k$	$\hat{\rho}_k$
1	0.502(0.02)	3.14(0.01)	0.90(0.01)	0.50(0.08)	3.10(0.10)	0.42(0.06)
2	0.252(0.02)	1.05(0.01)	0.89(0.01)	0.39(0.10)	0.67(0.18)	0.31(0.06)
3	0.241(0.03)	1.58(0.03)	0.79(0.03)	0.06(0.07)	0.09(0.40)	0.15(0.10)
4	0.003(0.00)	0.00(0.37)	0.36(0.23)	0.03(0.04)	0.03(0.39)	0.07(0.07)
5	0.001(0.00)	0.04(0.40)	0.17(0.17)	0.01(0.02)	0.01(0.39)	0.04(0.04)
6	0.001(0.00)	6.26(0.37)	0.09(0.10)	0.01(0.02)	0.01(0.39)	0.02(0.03)
7	0.00(0.00)	6.27(0.39)	0.04(0.07)	0.00(0.01)	0.01(0.39)	0.01(0.02)
8	0.00(0.00)	0.01(0.40)	0.02(0.04)	0.00(0.01)	0.01(0.39)	0.00(0.01)
9	0.00(0.00)	0.01(0.38)	0.01(0.02)	0.00(0.01)	0.01(0.38)	0.00(0.01)
10	0.00(0.00)	0.01(0.39)	0.01(0.01)	0.00(0.01)	0.01(0.38)	0.00(0.00)

Table 2. A comparison of clustering performance.

	DP mixture	K-means	Spherical K-means
Type 1	0.001	0.004	0.001
Type 2	0.052	0.071	0.071
Type 3	0.144	0.160	0.140

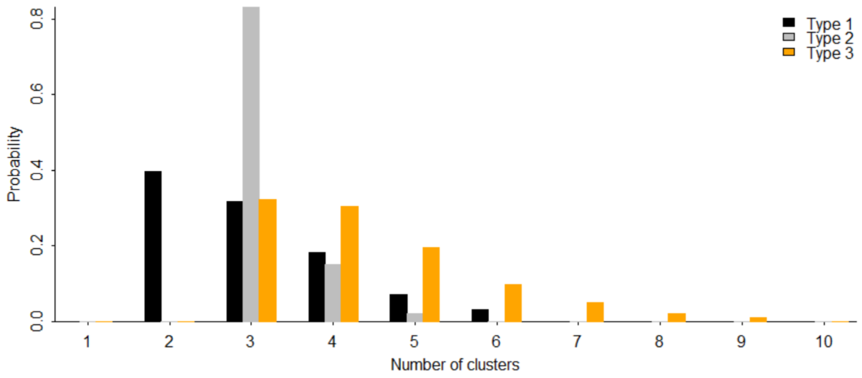
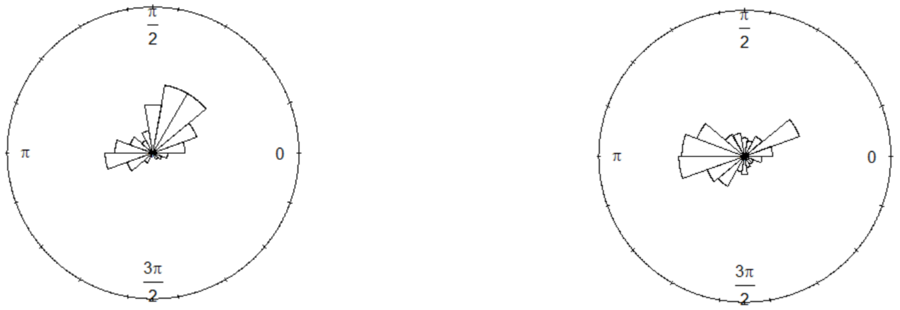


Figure 2. The posterior distribution of the number of clusters based on DP mixture circular models for simulation study.

4.2. A Monte Carlo simulation study

Aim of this section is to investigate important frequentist properties of estimators obtained using DP circular mixture approaches. We considered scenarios where approximation to DP circular mixture is achieved with a reasonably small C. In this regard, scenarios represented in panels (a) and (b) in Figure 3 are considered and following vM and wC mixture distributions are used to generate the Monte Carlo datasets for these scenarios respectively, for $i = 1, 2, ..., n$:

$$\begin{aligned} \theta_i &\sim 0.75 \text{ vM}(\pi/3, 5) + 0.25 \text{ vM}(\pi, 8), \\ \theta_i &\sim 0.20 \text{ wC}(\pi/6, 0.90) + 0.80 \text{ wC}(\pi, 0.70). \end{aligned}$$



(a) Scenario 1 (b) Scenario 2

Figure 3. Circular histograms.

For the analysis of these multiply-moded datasets, models in (3.1) and (3.2) are employed for datasets simulated from vM mixtures and wC mixtures respectively. In the application of model (3.1), priors $\mu_k \sim \text{vM}(0.1, 0.1)$, $\kappa_k \sim \text{Gamma}(1, 0.01)$, $k = 1, ..., C$ and $\alpha \sim \text{Uniform}(0.5, 10)$ are used while for the application of model (3.2), priors $\mu_k \sim \text{vM}(0.1, 0.1)$, $\rho_k \sim \text{Beta}(1, 1)$, $k = 1, ..., C$ and $\alpha \sim \text{Uniform}(0.5, 10)$ are used. Each scenario is repeated 250 times. Experiment is controlled for sample size (n is set at 100 for moderately large and at 500 for large samples as in biological studies in which circular data are most common). For each simulated dataset, first 10000 MCMC iterations are burnt in and following 10000 iterations are employed for posterior inference. Monte Carlo estimate of relative bias (MCRB), Monte Carlo standard error (MCSE), and average standard error estimate (ASE) are used to assess the frequentist properties of the estimators. Their formulas are given by MCRB $= (\frac{\bar{\hat{\beta}} - \beta}{\beta})$ where β is the true value of the particular parameter of interest, $\bar{\hat{\beta}} = \sum_{i=1}^{B} \hat{\beta}_i / B$, B is the number of Monte Carlo replications, and $\hat{\beta}_i$ is the Bayesian estimate of β for ith simulated dataset; MCSE $= \sqrt{[1/(B-1)] \sum_{i=1}^{B} (\hat{\beta}_i - \bar{\hat{\beta}})^2}$; ASE $= \sum_{i=1}^{B} \text{SE}(\hat{\beta}_i)/B$ where $\text{SE}(\hat{\beta}_i)$ is posterior standard error. MCSE and ASE measure the uncertainty in the estimators.

Tables 3 and 4 give the resulting estimators and their Monte Carlo properties for scenarios presented in panels (a) and (b) respectively. The true parameter values are given in the parentheses on the left hand sides of the tables and apply to right ones too. Accordingly, flexible DP mixture analysis well identified the clusters present in the data. In general, the estimators maintain good Monte Carlo properties with ignorable biases and relatively small standard errors. Moderately large samples may be needed for improved concentration estimation. Modest and comparable MCSE and ASE quantities are associated with lack of identifiability problems.

Table 3. Bayesian estimates (Est.), relative biases (MCRB), Monte Carlo standard errors (MCSE) and average posterior standard error estimates (ASE) for Scenario 1.

$n = 100$	Est.	MCRB	MCSE	ASE	$n = 500$	Est.	MCRB	MCSE	ASE
$\hat{\mu}_1$ (1.05)	1.05	0	0.06	0.06	$\hat{\mu}_1$	1.05	0	0.02	0.03
$\hat{\mu}_2$ (3.14)	3.14	0	0.08	0.08	$\hat{\mu}_2$	3.14	0	0.04	0.04
$\hat{\kappa}_1$ (5)	5.07	0.01	0.80	0.83	$\hat{\kappa}_1$	5.01	0	0.38	0.38
$\hat{\kappa}_2$ (8)	9.33	0.17	2.20	2.61	$\hat{\kappa}_2$	8.35	0.04	1.21	1.18
$\hat{p}_1$ (0.75)	0.74	-0.01	0.04	0.05	$\hat{p}_1$	0.75	0	0.01	0.02
$\hat{p}_2$ (0.25)	0.26	0.04	0.04	0.05	$\hat{p}_2$	0.25	0	0.01	0.02

Table 4. Bayesian estimates (Est.), relative biases (MCRB), Monte Carlo standard errors (MCSE) and average posterior standard error estimates (ASE) for Scenario 2.

$n = 100$	Est.	MCRB	MCSE	ASE	n=500	Est.	MCRB	MCSE	ASE
$\hat{\mu}_1$ (0.52)	0.52	0	0.04	0.05	$\hat{\mu}_1$	0.52	0	0.02	0.02
$\hat{\mu}_2$ (3.14)	3.14	0	0.06	0.06	$\hat{\mu}_2$	3.14	0	0.03	0.03
$\hat{\rho}_1$ (0.90)	0.88	-0.02	0.04	0.05	$\hat{\rho}_1$	0.90	0	0.02	0.02
$\hat{\rho}_2$ (0.70)	0.70	0	0.04	0.04	$\hat{\rho}_2$	0.70	0	0.02	0.02
$\hat{p}_1$ (0.20)	0.20	0	0.03	0.05	$\hat{p}_1$	0.20	0	0.01	0.02
$\hat{p}_2$ (0.80)	0.80	0	0.03	0.05	$\hat{p}_2$	0.80	0	0.01	0.02

4.3. Turtle data

We consider Gould's often studied turtle data that consists of the directions taken by 76 turtles after laying their eggs [29]. Earlier studies showed that the distribution of their orientations is bimodal. [18] considered a two-component finite mixture model of vM with different concentration parameters for its analysis and the results indicated that some turtles move into the direction of $\hat{\mu}_2 = 241.2^o$ while the majority ($\hat{p} = 84\%$) move to the direction of $\hat{\mu}_1 = 63.5^o$ where these are directions clockwise from North. Here we reanalyze the dataset to illustrate the use of doubly flexible DP mixture approach and to evaluate the contribution of the method relative to the standard finite mixture application. Circular histogram of the data that is given in Figure 4 in which some multi-modality strikes the eye.

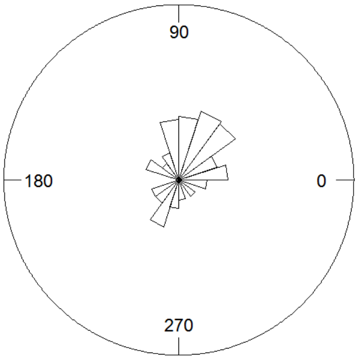


Figure 4. Circular histogram of turtle orientations.

Infinite mixture DP model is approximated by setting $C = 10$ and this is used as a start to identify the final number of clusters in the analysis. Baseline distribution for DP prior is set as $G_0 = \text{vM}(0.01, 7) \otimes \text{Gamma}(1, 0.01)$ for $k = 1, 2, ..., C$. Note that vague priors corresponding to μ_k's led to convergence problems due to the model complexity that increases with C and slightly more informative priors are hence more useful. Unknown concentration about the baseline distribution is assigned a Gamma prior, i.e. $\alpha \sim \text{Gamma}(2, 2)$. First 5000 MCMC iterations are burnt in and following 75000 iterations thinned by 5 steps are used for final posterior inference.

Posterior distribution of number of underlying directions (modes) is given in Figure 5. Accordingly, posterior probability for number of clusters is 0.39 and 0.33 for 2 (bimodal) and 3 (trimodal) respectively. The two probabilities are quite close suggesting the possibility of either bi or tri-modal directional preference. DP mixture-based approaches are particularly useful in such cases where it is hard to fix the number of clusters at a specific value prior to the analyis. For clustering purposes, the identified clusters based on DP mixture circular model are presented in Figure 6. Based on these results, it is seen that the two clusters for DP mixture vM model have been active in the Bayes 2.5% quantile and the Bayes 50% quantile for Gibbs sampler runs. Other clusters are active in the Bayes 97.5% quantile. Further to those, the DP mixture-based analysis has also given insight that about a few of the turtles under investigation preferred moving to the direction of other clusters based on the Bayes 97.5% quantile, a result that was not found before. According to these results, it appears that further research should be done in that area to better understand the phenomena in the general turtle population. The final result of the number of clusters based on the Bayes 50% quantile is estimated to be two. The rest of the posterior inference (parameter estimate and interquartile range (IQR)) by the DP circular mixture with the estimated number of clusters using the priors of the previous subsection as G_0 is given in Table 5. Accordingly, the analysis results suggest that the majority of the turtles move to mainly two different directions (about 83.5% of them to 63.426^o and 16.5% to 241.273^o). These results are consistent with the findings given in [18]. Hence, the DP mixture vM model demonstrates the superiority aspects of both determining the unknown number of mixture components and the identified clusters for turtle data.

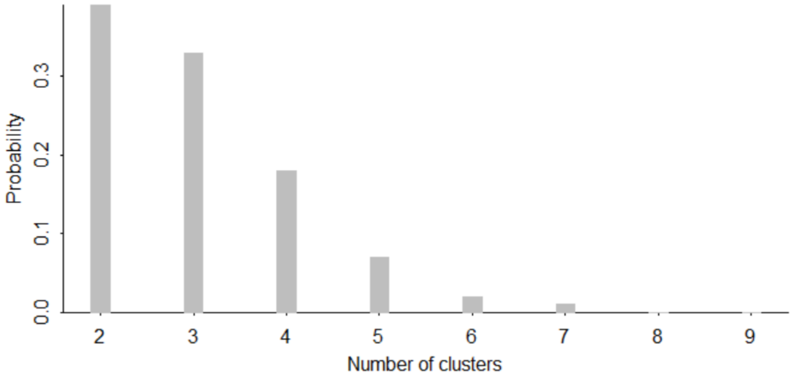


Figure 5. The posterior distribution of the number of clusters based on DP mixture vM model for turtle data.

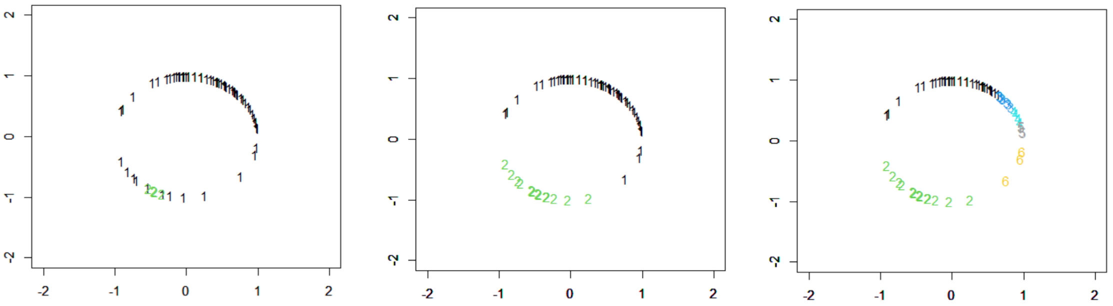

(a) Bayes 2.5% quantile (b) Bayes 50% quantile (c) Bayes 97.5% quantile

Figure 6. Bayesian quantiles and identified clusters for turtle data.

Table 5. Estimates and interquartile ranges of parameters for turtle data inference.

	p	μ_1	μ_2	κ_1	κ_2
Estimate	0.835	63.426^o	241.273^o	2.620	8.050
IQR	0.062	6.875^o	8.766^o	0.651	4.898

5. Concluding remarks

For the analysis of multi-modal circular data, the exact number of modes or number of groups present in the data is generally unknown and also there is no need to assume homogeneity of group specific concentration parameters. The DP mixture based approaches presented here along with use of stick breaking construction provided an insightful methodology for mixtures of the most common circular distributions. Our findings and comments in simulation sections provide guidance for practitioners in their data analysis and interpretation of the results. For large samples, initial C can be set at a larger value. Due to simulation costs, our Monte Carlo study was conducted only for cases where approximation to DP circular mixture is achieved with a reasonably small C. On the other hand, Section 4.1 presents empirical performance of the mixture method for a large C. Flexible DP mixture application on real data set has provided further insight into the phenomenon of animal orientations. This is particularly important in the age of increasing ecological challenges as it provides a holistic view of the world we live in.

The approach and considerations presented herein can be extended to mixture of skew circular distributions such as sine-skewed vM and wC as well as wide class of circular regression problems such as circular-circular and circular-linear.

Acknowledgment. We would like to thank the Editor and anonymous reviewers for their insightful comments which enabled us to improve the presentation of this paper. The first author also thanks to the Scientific and Technological Research Council of Turkey (TUBITAK, 2211-A) and Faculty Development Programme (ÖYP) for PhD scholarships.

References

[1] C. Agostinelli and U. Lund, *R package 'circular': Circular Statistics (version 0.4-93)*, 2017.

[2] S. Akesson, R. Klaassen, J. Holmgren, J.W. Fox and A. Hedenstrom, *Migration routes and strategies in a highly aerial migrant, the common swift Apus apus, revealed by light-level geolocators*, PLoS One **7** (7), 1-9, 2012.

[3] C.E. Antoniak, *Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems*, Ann. Statist. **2** (6), 1152-1174, 1974.

[4] S. Bhattacharya, and A. SenGupta, *Bayesian analysis of semiparametric linear-circular models*, J. Agric. Biol. Environ. Stat. **14** (1), 33-65, 2009.

[5] T. Ferguson, *A Bayesian analysis of some non-parametric problems*, Ann. Statist. **1** (2), 209-230, 1973.

[6] N.I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, 1993.

[7] M.D. Fraser, H. Yu-Sheng and J.J. Walker, *Identifiability of finite mixtures of von Mises distributions*, Ann. Statist. **9** (5), 1130-1131, 1981.

[8] R. Gatto and S.R. Jammalamadaka, *The generalized von Mises distribution*, Stat. Methodol. **4** (3), 341-353, 2007.

[9] K. Ghosh, S.R. Jammalamadaka and R. Tiwari, *Semiparametric Bayesian techniques for problems in circular data*, J. Appl. Stat. **30** (2), 145-161, 2003.

[10] T. Guilford, R. Freeman, D. Boyle, B. Dean, H. Kirk, R. Phillips and C. Perrins, *A dispersive migration in the Atlantic puffin and its implications for migratory navigation*, PLoS One **6** (7), 1-8, 2011.

[11] E.J. Gumbel, *Applications of the circular normal distribution*, J. Amer. Statist. Assoc. **49** (266), 267-297, 1954.

[12] H. Holzmann, A. Munk and B. Stratmann, *Identifiability of finite mixtures - with applications to circular distributions*, Sankhya **66** (3), 440-449, 2004.

[13] K. Hornik, I. Feinerer, M. Kober and C. Buchta, *Spherical k-means clustering*, J. Stat. Softw. **50** (10), 1-22, 2012.

[14] H. Ishwaran and L.F. James, *Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information*, J. Comput. Graph. Statist. **11** (3), 508-532, 2002.

[15] S.R. Jammalamadaka, B. Wainwright and Q. Jin, *Functional clustering on a circle using von Mises mixtures*, J. Stat. Theory Pract. **15** (2), 1-17, 2021.

[16] T.A. Jones and W.R. James, *Analysis of bimodal orientation data*, J. Int. Assoc. Math. Geol. **1** (2), 129-135, 1969.

[17] S. Kim and A. SenGupta, *A three-parameter generalized von Mises distribution*, Statist. Papers **54** (3), 685-693, 2013.

[18] K.V. Mardia, *Statistics of directional data*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **37** (3), 349–393, 1975.

[19] K.V. Mardia and T.W. Sutton, *On the modes of a mixture of two von Mises distributions*, Biometrika **62** (3), 699-701, 1975.

[20] R.M. Neal, *Markov chain sampling methods for Dirichlet process mixture models*, J. Comput. Graph. Statist. **9** (2), 249-265, 2000.

[21] G. Nunez-Antonio, M.C. Ausin and M.P. Wiper, *Bayesian nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions*, J. Agric. Biol. Environ. Stat. **20** (1), 47-64, 2015.

[22] G. Nunez-Antonio, M. Mendoza, A. Contreras-Cristan, E. Gutierrez-Pena and E. Mendoza, *Bayesian nonparametric inference for the overlap of daily animal activity patterns*, Environ. Ecol. Stat. **25** (4), 471-494, 2018.

[23] A. Ożarowska, M. Ilieva, P. Zehtindjiev, S. Akesson and K. Muś, *A new approach to evaluate multimodal orientation behaviour of migratory passerine birds recorded in circular orientation cages*, J. Exp. Biol. **216** (21), 4038-4046, 2013.

[24] A. Pewsey, *The wrapped stable family of distributions as a flexible model for circular data*, Comput. Statist. Data Anal. **52** (3), 1516-1523, 2008.

[25] J.R. Pcyke, *Some tests for uniformity of circular distributions powerful against multimodal alternatives*, Canad. J. Statist. **38** (1), 80-96, 2010.

[26] A. SenGupta and M. Roy, *Clustering on the torus*, J. Stat. Theory Pract. **15** (3), 1-18, 2021.

[27] A. SenGupta, M. Roy and A.K. Chattopadhyay, *Model-based clustering for cylindrical data*, in: I. Ghosh, N. Balakrishnan, H.K.T. Ng (ed.) Advances in Statistics - Theory and Applications, Emerging Topics in Statistics and Biostatistics, 347-363, Springer, Cham, 2021.

[28] J. Sethuraman, *A constructive definition of Dirichlet priors*, Statist. Sinica **4** (2), 639-650, 1994.

[29] M.A. Stephens, *Techniques for directional data*, Technical Report 150, Stanford University, Department of Statistics, 1969.

[30] S. Sturtz, U. Ligges and A. Gelman, *R2WinBUGS: A Package for running WinBUGS from R*, J. Stat. Softw. **12** (3), 1-16, 2005.

[31] E.A. Yfantis and L.E. Borgman, *An extension of the von Mises distribution*, Comm. Statist. Theory Methods **11** (15), 1695-1706, 1982.

Appendix

Posterior computation for DP mixture vM model

Let $K_1^*,...,K_m^*$ be the current m unique values of $\boldsymbol{K}$. Conditional distribution $\boldsymbol{\varphi}|\boldsymbol{K},\boldsymbol{\theta}$ can be decomposed as $\boldsymbol{\mu}|\boldsymbol{\kappa},\boldsymbol{K},\boldsymbol{\theta}$ and $\boldsymbol{\kappa}|\boldsymbol{\mu},\boldsymbol{K},\boldsymbol{\theta}$. Then, in each iteration of Gibbs sampler, we draw samples from the full conditional densities listed below.

Conditional distribution for $\boldsymbol{\mu}$: for each $j \in K_1^*,...,K_m^*$, draw

$$\mu_j|\boldsymbol{\kappa},\boldsymbol{K},\boldsymbol{\theta} \propto \exp(\kappa_0 \cos(\mu_j - \mu_0) + \sum_{i:K_i=j} \kappa_j \cos(\theta_i - \mu_j)).$$

In OpenBUGS, block-hybrid sampling algorithm is performed to obtain random samples from $\mu_j|\boldsymbol{\kappa},\boldsymbol{K},\boldsymbol{\theta}$. Also for each $j \in \boldsymbol{K} - K_1^*,...,K_m^*$, independently simulate $\mu_j \sim \mathrm{vM}(\mu_0,\kappa_0)$.

Conditional distribution for $\boldsymbol{\kappa}$: for each $j \in K_1^*,...,K_m^*$, draw

$$\kappa_j|\boldsymbol{\mu},\boldsymbol{K},\boldsymbol{\theta} \propto \frac{\kappa_j^{b_0-1}}{I_0(\kappa_j)^{n_j}} \exp(\sum_{i:K_i=j} \kappa_j \cos(\theta_i - \mu_j) - a_0\kappa_j),$$

where $n_j = \#\{i : K_i = j\}$. Slice sampling algorithm is used to obtain random samples from the full conditional distribution of κ_j. Also for each $j \in \boldsymbol{K} - K_1^*,...,K_m^*$, independently simulate $\kappa_j \sim \mathrm{Gamma}(a_0,b_0)$.

Conditional distribution for $\boldsymbol{K}$:

$$(K_i|\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{\kappa},\boldsymbol{\theta}) \sim \sum_{k=1}^{C} p_{k,i}\ I_k(.),\ i = 1,...,n$$

where

$$(p_{1,i},...,p_{C,i}) \propto \frac{p_1}{I_0(\kappa_1)} \exp(\kappa_1(\cos(\theta_i - \mu_1)),..., \frac{p_C}{I_0(\kappa_C)} \exp(\kappa_C(\cos(\theta_i - \mu_C)).$$

Discrete slice sampling algorithm is used to obtain random samples from the full conditional distribution of $\boldsymbol{K}$

Conditional distribution for $\boldsymbol{p}$:

$$p_1 = q_1^* \text{ and } \ p_k = (1-q_1^*)(1-q_2^*)...(1-q_{k-1}^*)q_k^*, \quad k = 2,...,C-1$$

where

$$q_k^* \sim \mathrm{Beta}(1 + n_k, \alpha + \sum_{l=k+1}^{C} n_l),\ k = 1,...,C-1$$

where $n_k = \#\{i : K_i = k\}$, that is, n_k saves the number of K_i values which set to k

Conditional distribution for α:

$$\alpha|\boldsymbol{p} \sim \text{Gamma}(C + v_1 - 1, v_2 - \sum_{k=1}^{C-1} \log(1 - q_k^*),$$

where q_k^* are same values in the simulation of $\boldsymbol{p}$.

Posterior computation for DP mixture wC model

Let $K_1^*, ..., K_m^*$ be the current m unique values of $\boldsymbol{K}$. Conditional distribution $\boldsymbol{\varphi}|\boldsymbol{K},\boldsymbol{\theta}$ can be decomposed as $\boldsymbol{\mu}|\boldsymbol{\rho},\boldsymbol{K},\boldsymbol{\theta}$ and $\boldsymbol{\rho}|\boldsymbol{\mu},\boldsymbol{K},\boldsymbol{\theta}$. Then, in each iteration of Gibbs sampler, we draw samples from the full conditional densities listed below.

Conditional for $\boldsymbol{\mu}$: for each $j \in K_1^*, ..., K_m^*$, draw

$$\mu_j|\boldsymbol{\rho},\boldsymbol{K},\boldsymbol{\theta} \propto \exp(\kappa_0 \cos(\mu_j - \mu_0) + \sum_{i:K_i=j} \log(\frac{1}{1 + \rho_j^2 - 2\rho_j \cos(\theta_i - \mu_j)})),$$

it can be used Taylor expansion of $\log(1/1 + x))$ with ignored high order terms, we repeatedly write the full conditional distributions for $\boldsymbol{\mu}$ as follows

$$\propto \exp(\kappa_0 \cos(\mu_j - \mu_0) + \sum_{i:K_i=j} 2\rho_j \cos(\theta_i - \mu_j) - \rho_j^2)),$$

and then, the block-hybrid sampling algorithm is performed to obtain random samples from the full conditional distribution of μ_j. Also for each $j \in \boldsymbol{K} - K_1^*, ..., K_m^*$, independently simulate $\mu_j \sim \text{vM}(\mu_0, \kappa_0)$.

Conditional distribution for $\boldsymbol{\rho}$: for each $j \in K_1^*, ..., K_m^*$, draw

$$\rho_j|\boldsymbol{\mu},\boldsymbol{K},\boldsymbol{\theta} \propto \rho_j^{a_0-1}(1 - \rho_j)^{b_0-1} \prod_{i:K_i=j} \frac{1 - \rho_j^2}{1 + \rho_j^2 - 2\rho_j \cos(\theta_i - \mu_j)}.$$

Slice sampling algorithm is performed to obtain random samples from $\rho_j|\boldsymbol{\mu},\boldsymbol{K},\boldsymbol{\theta}$. Also for each $j \in \boldsymbol{K} - K_1^*, ..., K_m^*$, independently simulate $\rho_j \sim \text{Beta}(a_0, b_0)$.

Conditional distribution for $\boldsymbol{K}$:

$$(K_i|\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\theta}) \sim \sum_{k=1}^{C} p_{k,i}\ I_k(.),\ i = 1, ..., n$$

where

$$(p_{1,i}, ..., p_{C,i}) \propto p_1 \frac{1 - \rho_1^2}{1 + \rho_1^2 - 2\rho_1 cos(\theta_i - \mu_1)}, ..., p_C \frac{1 - \rho_C^2}{1 + \rho_1^2 - 2\rho_C cos(\theta_i - \mu_C)}.$$

The discrete slice sampling algorithm is used to obtain random samples from the full conditional distribution of $\boldsymbol{K}$. Other full conditional distributions are similar manner with DP mixture vM model.