International Journal of Mathematics and
Computer Science, **19**(2024), no. 4, 1213–1221

Densities Estimation for Bivariate Circular Data Under Jones-Pewsey Circular Kernel

Huda Kareem Nasser
Samira Faisal Abushilah

Department of Mathematics
Faculty of Education for Girls
University of Kufa
Najaf, Iraq

email: hodak.alkhafaji@student.uokufa.edu.iq
Sameerah.hathoot@uokufa.edu.iq

(Received March 27, 2024, Accepted May 10, 2024,
Published June 1, 2024)

Abstract

Owing to the fact that a distribution is more informative than classical representations, we explore an estimation of the densities of bivariate directional data using a kernel density estimation technique. Jones-Pewsey distribution is used as a kernel function for estimating the densities for bivariate angular data points which are frequently encountered in many fields such as medicine, biology, astronomy, geography, and ecology. Moreover, using R software, we perform a simulation study to evaluate the methodology that we have proposed under observed bivariate angular data set which are generated from von-Mises circular distribution under different modes, different sample size, as well as different random smoothing parameters.

1 Introduction

Angular data which can be shown as points on the circle and can be represented by an angle $\theta \in [0, 2\pi)$ are periodic (i.e., $\theta = \theta + 2n\pi$, where $n \in Z$).

Keywords and phrases: Bivariate circular data, kernel density estimation, Jones-pewsey distribution, Von-mises distribution.
AMS (MOS) Subject Classifications: 62Hxx, 62H11.
ISSN 1814-0432, 2024, http://ijmcs.future-in-tech.net

To deal with this kind of data which is distinct from traditional linear data, we need new statistical techniques which are relatively new and still under development [1]. In several instances, there is an occurrence of directional data, prompting our interest in determining the density function of a number of angle random variables [2]. One of the most popular methods for estimating the underlying probability density function of a dataset is kernel density estimation (KDE). As a nonparametric density estimation, KDE will automatically determine the density's shape based on the data [3].

In the literature, there are a few contributions on data lying on the sphere or on the unit circle. For spherical data, Hall, Watson, and Cabrera [4] used circular derivatives to study KDE, while for data lying on the unit circle Fisher [5] was the first to propose a KDE by modifing the method of Silverman [6] which was usually used for linear data where a quartic kernel function has been used which is not a circular kernel. Even so, until 2011 there was nothing presented for kernel on the torus.

For a p-dimensional torus, Di Marzio, Panzera, and Taylor [2] extended the circular KDE under the hypothesis that the parameters κ_1 and κ_2 for the kernel functions are equate. In 2012, Taylor et al. [7] proposed a method for analyzing the structure of protein by estimating tailed probability based on the estimated density of amino acids. To accurately estimate density on a torus, a bivariate kernel function with various smoothing and concentration parameters is required [8]. Therefore, for bivariate angular data and under the assumption ($\kappa_1 = \kappa_2$), we seek to develop a methodology to estimate the density function.

On fitting directional distributions, for example von-Mises and wrapped normal, there has been a few contributions in recent years. In 2014, Hornik et al. [9] developed movMF, an R package that utilizes the expectation maximization method to accurately determine finite mixes of von Mises-Fisher distributions. In 2017, Chakraborty et al. [10] presented BAMBI, an R package that utilizes Bayesian techniques and finite mixture models to model univariate and bivariate angular data for circular distributions like von Mises and wrapped normal. Even though the packages mentioned before serve as a good step for fitting bivariate von-Mises model and normal wrapped model, we seek to estimate the density function for bivariate directional data on the torus without making the hypothesis about the distribution of that data points. In 2019, Abushilah [11] proposed a method for estimating the density function for angular data which is bivariate by using KDE technique under von-Mises kernel function. In 2020, Abushilah and et al. [12] presented a methodology to cluster some groups of angles and they applied this methodology to cluster dihedral angles which belong to some amino acids. Therefore, in this paper, we suggest a methodology which can be applied to estimate the densities of bivariate circular data (BCD) $(\phi_i, \psi_i), i = 1, 2, ..., m$ using Jones-Pewsey circular kernel with kernel density estimation technique under different concentration parameters. In Section 2, we present the algorithm that we have suggested to estimate the density function for bivariate angular data points. Moreover, we apply the methodology which we have shown in Section 2 to the data which will be generated from von-Mises distribution to visualise its density. Finally, we present a conclusion to our paper.

2 Densities Estimation for BCD

KDE is a non-parametric technique which can be used to estimate the pdf of a random variable [13]. Let us start with a few topics to get an idea of how KDE is utilized. For a random sample which is taken from a population with unknown density, the univariate kernel estimate of the original one is given by [14]:

$$\widehat{f}(x) = \frac{1}{m}\sum_{i=1}^{m} K\big(\frac{x - x_i}{B}\big), \tag{2.1}$$

where B is the free parameter in the kernel function $K(\cdot)$ (also called the bandwidth or smoothing parameter) such that:

- $\int_{-\infty}^{\infty} K(\cdot)dx = 1.$
- $K(\cdot)$ is nonnegative.
- $K(\cdot)$ is bounded.

In this paper, we propose an algorithm to estimate the densities for bivariate directional data points. In this approach, we use the KDE technique with the Jones-Pewsey circular kernel under different concentration parameters.

The Jones-Pewsey distribution $JP(\mu, \kappa, \varepsilon)$, proposed by Jones and Pewsey [15], is a unimodal distribution which is symmetric with the following pdf

$$f(\theta; \mu, \kappa, \varepsilon) = \frac{[cosh(\kappa\varepsilon) + sinh(\kappa\varepsilon)cos(\theta - \mu)]^{1/\varepsilon}}{2\pi P_{1/\varepsilon}\; cosh(\kappa\varepsilon)}, 0 \leq \theta < 2\pi, \tag{2.2}$$

where the parameter μ is the mean direction, κ is the concentration parameter, ε is the shape parameter, and $P_{1/\varepsilon}(\cdot)$ is the associated Legendre function of the first kind, degree $1/\varepsilon$ and order zero. The three-parameter Jones-Pewsey distribution is useful for modeling and tracking circular data.

For bivariate circular data, $\{(\phi_1, \psi_1), (\phi_2, \psi_2), \ldots, (\phi_m, \psi_m)\}$, a kernel density estimate $\widehat{f}(\phi, \psi)$ can be constructed using the kernel function in Equation (2.2) with parameters μ and κ where the parameter κ corresponds to the smoothing parameter. The KDE for angular data points is constructed using the following formula:

$$\widehat{g}(\phi, \psi) = \frac{1}{m} \sum_{i=1}^{n} \left[\mathrm{K}_1(\phi - \phi_i; \mu_1, \kappa_1, \varepsilon_1) \mathrm{K}_2(\psi - \psi_i; \mu_2, \kappa_2, \varepsilon_2) \right], \tag{2.3}$$

where (κ_1, κ_2) are the smoothing parameters (bandwidth) and $\mathrm{K}_j(\cdot), j = 1, 2$ are the kernel functions, Equation (2.2). A multivariate kernel in Equation (2.3) is used because the data is angular and bivariate such that for each observation there is a particular kernel, where the parameters are different. We show the suggested procedure in Algorithm 1.

Algorithm 1 Estimating densities for bivariate circular data using Jones-Pewsey distribution.

1. Let $\{(\phi_1, \psi_1), ..., (\phi_m, \psi_m)\}$ be a bivariate angular data with an unknown density, say $\widehat{g}(\phi, \psi)$.
2. Given $m_1 \in Z$ as the number of grid.
3. Generate $\{(r_x, r_y)\}$ over the range $(-\pi, \pi)$ such that each entry is given by
$$r_i = \frac{-\pi(m_1 - 1) + (c - 1)2\pi}{(m_1 - 1)}, c = 1, ..., m_1.$$
4. Select B_1 and B_2 as the random bandwidth.
5. Calculate $\triangle_x = r_x - \phi$ & $\triangle_y = r_y - \psi$.
6. For each combination of the sequences $\{\phi_1, ..., \phi_n\}$ and $\{\psi_1, ..., \psi_n\}$ calculate $\widehat{g}(\phi, \psi)$ using Equation (2.3) by setting $\phi = r_x$ and $\psi = r_y$.

3 Simulation Study

The procedure which is presented in Section 2 is applied to the generated data which are simulated from different models from the BvM(μ, Σ), Equation

(3.5), to highlight the joint pdf for the generated data.

$$f(\theta_1, \theta_2) = C_{sin} \exp\Big\{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \\ \sin(\theta_2 - \mu_2)\Big\}, 0 \leq \phi, \psi < 2\pi, \tag{3.4}$$

$$C_{sin}^{-1} = 4\pi^2 \sum_{r=0}^{r} \binom{2r}{r} \left(\frac{\lambda^2}{4\kappa_1\kappa_2}\right)^r I_r(\kappa_1) I_r(\kappa_2), \tag{3.5}$$

where $0 \leq \mu_1, \mu_2 < 2\pi$ and $\kappa_1, \kappa_2 \geq 0$ are the parameters of the distribution, the first one represents the circular mean while the second one represents the concentration, respectively while $I_r(\kappa)$ represents the first kind modified Bessel function of order zero which is defined by

$$I_r(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(r\theta) \exp\left\{\kappa \cos(\theta - \mu)\right\} d\theta, r = 0, 1, 2, \ldots \tag{3.6}$$

Different models are generated which are summarized in the following steps:

- Model 1: angular data of different sizes $m \in \{300, 500, 800, 1000\}$ are generated from BvM(μ, Σ), where $\mu = (0, 0)^T$ and $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
- Model 2: angular data of different sizes $m \in \{300, 500, 800, 1000\}$ are generated from BvM(μ, Σ), where $\mu = (2, 2)^T$ and $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$.
- Model 3: angular data of different sizes $m \in \{300, 500, 800, 1000\}$ are generated from BvM(μ, Σ), where $\mu = (0.3, 0.3)^T$ and $\Sigma = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$.

Some of the evaluation graphs produced are shown in Figures 1 and 2. Figure 1 shows the square root of the probability density function for the simulated circular data while Figure 2 highlights the contour plot for the circular data points under different smoothing parameters. In these Figures, we can see the following:

1. The shape of the joint probability density function for angles (ϕ, ψ) is different from one model to another. This suggests that the joint pdf depends on the parameters of the model.
2. The free parameters in the kernel functions has a strong impact on the shape of the joint pdf of the circular data.

3. The contour plot for joint pdf is different as well from one model to another (see Figure 2).

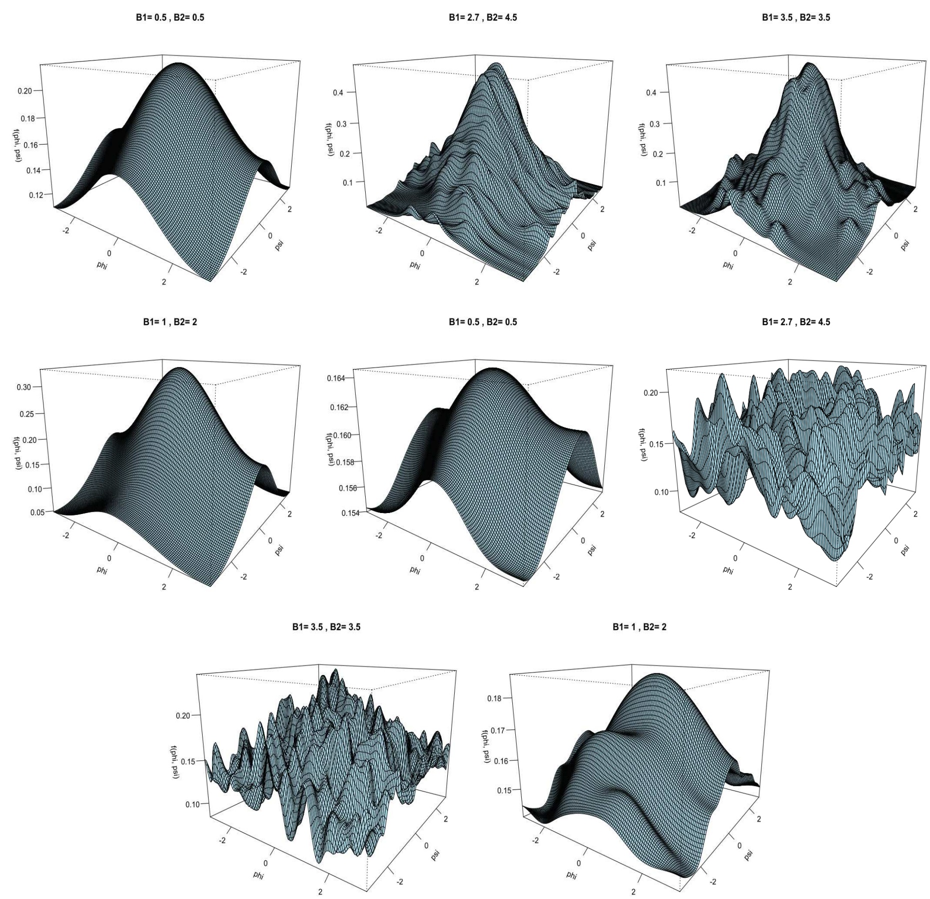

Figure 1: The joint pdf for Models 1, 3 under different smoothing parameters.

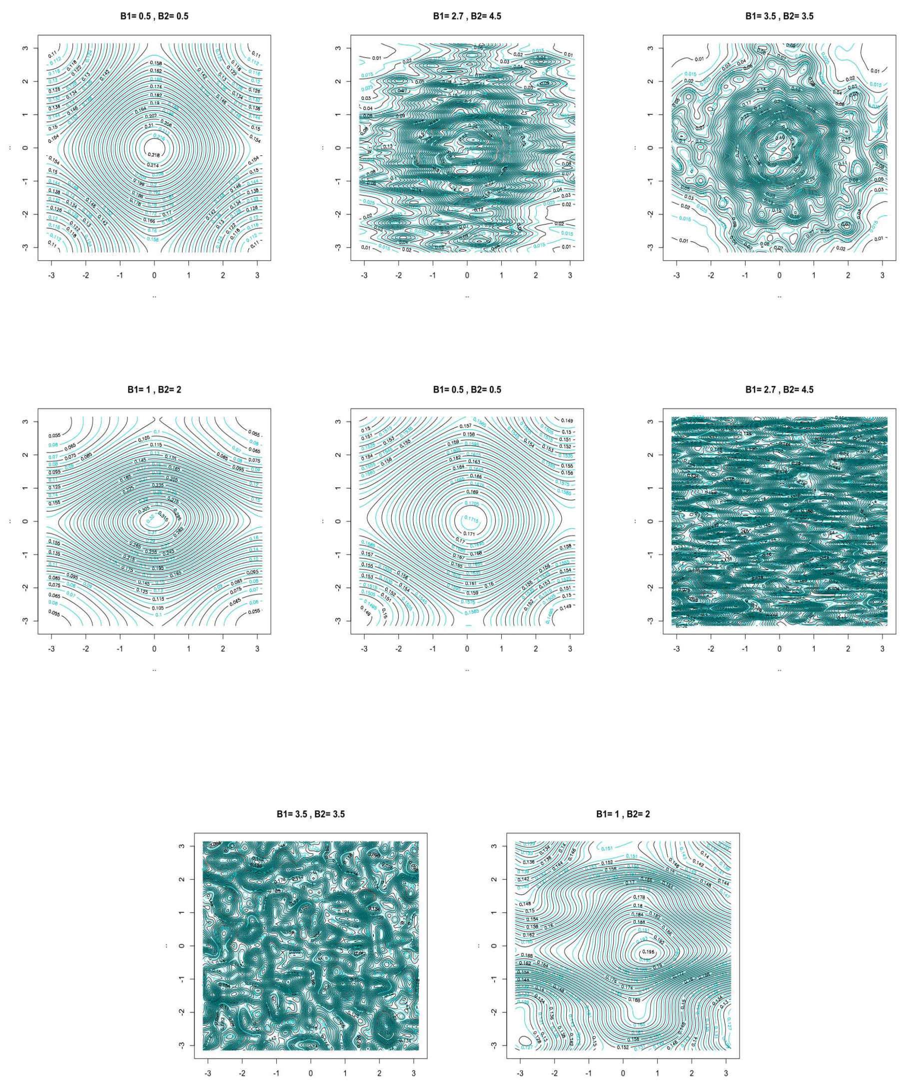

Figure 2: Contour plot for Models 1, 3 with different smoothing parameters.

4 Conclusion

In this paper, a procedure for estimating the joint pdf for bivariate angular data was suggested using kernel a density estimation technique with a Jones-Pewsey distribution as a kernel function. Then the proposed algorithm was applied to generated data which have been taken from von-Mises circular distribution to evaluate the performance of the methodology that we have suggested under different modes, different sample size, as well as different random smoothing parameters using R software. The results of the simulation indicated that the shape of the joint pdf for the bivariate directional data is different from one model to another showing that the joint pdf depends on the parameters of the model. The free parameters in the kernel functions has a strong impact on the shape of the joint pdf of the circular data.

References

[1] K.V. Mardia, P.E. Jupp, Directional Statistics, John Wiley & Sons, 2000.

[2] M. Di Marzio, A. Panzera, C.C. Taylor, Kernel density estimation on the torus, Journal of Statistical Planning and Inference, **141,** no. 6, (2011), 2156–2173.

[3] E. Parzen, On estimation of a probability density function and mode, The Annals of Mathematical Statistics, **33,** no. 3, (1962), 1065–1076.

[4] P. Hall, G. Watson, J. Cabrera, Kernel density estimation with spherical data, Biometrika, **74,** no. 4, (1987), 751–762.

[5] N. Fisher, Smoothing a sample of circular data, Journal of Structural Geology, **11,** no. 6, (1989), 775–778.

[6] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, 1986.

[7] C.C. Taylor, K.V. Mardia, M. Di Marzio, A. Panzera, Validating protein structure using kernel density estimates, Journal of Applied Statistics, **39,** no. 11, (2012), 2379–2388.

[8] H. Sasaki, Y.K. Noh, M. Sugiyama, Direct density-derivative estimation and its application in kl-divergence approximation, in Artificial Intelligence and Statistics, (2015), 809–818.

[9] K. Hornik, B. Grun, movMF: an R package for fitting mixtures of von mises-fisher distributions, Journal of Statistical Software, **58,** no. 10, (2014), 1–31.

[10] S. Chakraborty, S.W. Wong, BAMBI: An R package for fitting bivariate angular mixture models, arXiv:1708.07804, (2017).

[11] S.F.H. Abushilah, Clustering methodology for bivariate circular data with application to protein dihedral angles, PhD thesis, University of Leeds, (2019).

[12] S.F. Abushilah, C.C. Taylor, A. Gusnanto, Geometry-based distance for clustering amino acids, Journal of Applied Statistics, **47**, no. 7, (2020), 1235–1250.

[13] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman & Hall/CRC Press, 1995.

[14] A. Gramacki, Nonparametric kernel density estimation and its computational aspects, Springer, 2018.

[15] M. Jones, A. Pewsey, A family of symmetric distributions on the circle, Journal of the American Statistical Association, **100,** no. 472, (2005), 1422–1428.