International Journal of Mathematics and
Computer Science, **20**(2025), no. 1, 23–31
DOI: https://doi.org/10.69793/ijmcs/01.2025/huda

Smoothing Parameters Selection for Samples from Bivariate Circular Distributions

Huda Kareem Nasser, Samira Faisal Abushilah

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

email: hudakarem@utq.edu.iq,
Sameerah.hathoot@uokufa.edu.iq

(Received May 15, 2024, Revised June 19, 2024,
Accepted June 21, 2024, Published June 28, 2024)

Abstract

Kernel density estimates for bivariate circular data are efficient non-parametric estimation methods incorporating free smoothing parameters that significantly influence the estimation process's results. In this paper, we focus our attention on selecting the optimal bandwidth for bivariate circular data from the von Mises distribution using cross-validation and a nonlinear minimized method using circular packages in R software.

1 Introduction

Circular data, measured in degrees or radians and represented as a point on a circle, is fundamentally different from linear data due to its periodic character ($0° = 360°$) [1]. Circular datasets are often seen in several fields including biological sciences, bioinformatics, meteorology, and geography [2]. Special techniques must be used for angular data and, when construct these techniques, we must take into consideration the nature of this type of data.

Keywords and phrases: Bivariate circular data, Smoothing parameters, Kernel density estimation, Cross-validation, Maximum likelihood estimation, Jones-Pewsey Distribution.
AMS (MOS) Subject Classifications: 62Hxx , 62H11.
ISSN 1814-0432, 2025, http://ijmcs.future-in-tech.net

In the literature, for the univariate angular data, the smoothing parameter (bandwidth) selection received little attention. In 2008, Taylor [3] proposed a plug-in rule approach to choose the optimal smoothing parameters for univariate angular density estimation. In 2011, Di marzio et al. [4] used the bootstrap approach to consider a procedure which can be used to select the bandwidth for circular data. In 2012, Oliveira et al. [5] presented a new plugin method to select the smoothing parameters for the kernel density estimation which is evaluated using simulated data. In 2019, for multivariate angular data points, Abushilah [6] used von-Mises distribution as a kernel function to select optimal bandwidth for amino acids from 500 proteins. In 2023, Zámečník et al. [2] presented common bandwidth selection methods focusing on variable bandwidth selection where they used simulations and real datasets to evaluate and compare these methods and highlighting their potential advantages.
In this paper, we propose a methodology to determine smoothing parameters for bivariate circular data using Cross Validation to enhance model accuracy, MLE to ensure that estimates converge to true parameter values, and the Jones-Pewsey distribution, characterized by its parameters that can be modified to suit different types of circular data in scientific fields. We use optimization to improve performance and calculate kernel density within the optimal bandwidth.

2 Selection of Bandwidth for Bivariate Angular Data

Let $(\varphi_1, \psi_1), \in [-\pi, \pi], i = 1, 2, \ldots,\ n$ be an angular data from a population with an unknown density $f(\varphi, \psi)$. The definition of the circular kernel density estimation $\hat{f}(\varphi, \psi)$ of function $f(\varphi, \psi)$ is as follows:

$$\hat{f}\,(\varphi, \psi) = \frac{1}{n} \sum_{i=1}^{n} K(\frac{\varphi - \varphi_i}{h_1}) K(\frac{\psi - \psi_i}{h_2}) \tag{2.1}$$

Leave one-out cross-validation (LOOCV) is a cross-validation technique. In this method, the dataset is partitioned in a manner where one observation is designated as the validation set, while the remaining observations are allocated as the training set. Later, the above will be applied in our work. The value of bandwidth, which is the concentration parameter for the Jones-Pewsey distribution, should be taken into consideration for every given value of $\kappa = (\kappa_1, \kappa_2) > 0$, the dataset is divided into two partitions; namely, $\mathscr{B}_1$ and $\mathscr{B}_1$. The initial partition, denoted as $\mathscr{B}_1$ and referred to as the test partition, comprises of a single perpendicular pair of angles $\mathscr{B}_1 = (\varphi_i, \psi_i)\,,\ i = 1, 2, \ldots,\ n$. Others are contained in the second partition $\mathscr{B}_2 = (\varphi_j, \psi_j)\,,\ j = 1, 2, \ldots,\ n,\ i \neq j$. When working with B1 partition data, the kernel density estimation is determined by using this formula:

$$\hat{f}^{(-i)}(\varphi_j, \psi_j) = \frac{1}{n-1}\left[\sum_{j \neq i} K(\triangle_1; \mu, \kappa_1)\ \ K(\triangle_2; \mu, \kappa_2)\right] \tag{2.2}$$

where $\triangle_1 = \varphi_i - \varphi_j,\ \triangle_2 = \psi_i - \psi_j, \mu = 0$, and $K(\cdot)$ As a kernel function represents the Jones-Pewsey distribution, which is defined by [8]:

$$f(\varphi; \mu; \kappa) = \frac{[\cosh(\kappa\lambda) + \sinh(\kappa\lambda)\cos(\varphi - \mu)]^{\frac{1}{\lambda}}}{2\pi P_{\frac{1}{\lambda}} \cosh(\kappa\lambda)}, 0 \leq \varphi \leq 2\pi \tag{2.3}$$

where μ is the mean direction $(0 \leq \mu \leq 2\pi)\,, \kappa$ a concentration parameter $(\kappa \geq 0)$, λ is a shape parameter $(-\infty < \lambda < \infty)$, and $P_{\frac{1}{\lambda}}(z)$ is the associated Legendre function of the first kind of degree $\frac{1}{\lambda}$ and order 0. Compute the likelihood function utilizing the circular kernel function stated in Eq. 2.3 with the given formula:

$$L(\kappa; (\varphi_1, \psi_1), (\varphi_2, \psi_2), \ldots, (\varphi_n, \psi_n)) = \prod_{i=1}^{n} \hat{f}^{(-i)}(\varphi_j, \psi_j) \tag{2.4}$$

Then calculate the likelihood function's logarithm using Eq. 2.4 to get

$$L(\kappa; (\varphi_1, \psi_1), (\varphi_2, \psi_2), \ldots, (\varphi_n, \psi_n)) = \sum_{i=1}^{n} \log\left(\hat{f}^{(-i)}(\varphi_j, \psi_j)\right) \tag{2.5}$$

The next step is to apply the optimization method to our proposed methodology in order to boost its performance. We have utilized the Newton-Raphson method in conjunction with a non-linear minimization technique (nlm-technique) for optimization purposes [9]. A good approximation for the real-valued, continuous, and differentiable function can be quickly found using this method. The Raphson technique for solving equations with one variable is executed as follows: The function $f(y)$ is assumed to have a root close to the equation $y = y_n$, According to Newton's method, a more accurate estimate of the root is

$$y_{n+1} = y_n - \frac{f(y_n)}{\acute{f}(y_n)}, \tag{2.6}$$

such that at the point $y = y_n$, the slope for the line that is tangent to the graph $f(y)$ is $\acute{f}(y_n)$. Setting $y_n = y_{n+1}$ repeatedly iteratively improves the approximation and achieves the target accuracy according to Eq. 2.6 until y_{n+1} approaches y_n. Using a Newton-Raphson approach, as shown above, the function nlm (non-linear minimization) in the R statistical package [10] minimizes the function $f(y)$. IN Algorithm 1, we present the approach that we suggest in this section for improving the performance given by the (nlm-BLCV). Algorithm 1 (nlm-BLCV) utilizes the non-linear minimizing function nlm($f(\kappa_1, \kappa_2)$). Here, f represents the log-likelihood-cross-validation function that we want to minimize, while (κ_1, κ_2) denotes the beginning values. For bivariate circular data $\{(\varphi_1, \psi_1), (\varphi_2, \psi_2), \ldots, (\varphi_n, \psi_n)\}$, the optimal smoothing parameters are the ones that solve the nlm $(f(\kappa_1, \kappa_2))$.

3 Evaluation of the Performance

The process described in Section 2 applies to the data generated, which are simulated from two models from the $BvM(\mu, \kappa)$ distribution. The next steps summarize the generation of two models:

- Model 1: circular data of varying sizes m are produced by $BvM(\mu, \kappa)$

$$\text{where } \mu = (0, 0)^T, \kappa = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Model 2: circular data of varying sizes m are produced by $BvM(\mu, \kappa)$

$$\text{where } \mu = (0.3,\ 3.5)^T,\ \kappa = \begin{pmatrix} 1.2 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

In Table 1, we present the results of the joint probability density function for the two models above, obtained using the technique described in Section 2.

The optimal bandwidth was calculated. We have observed the following:

1. The parameter influenced by the sample size, but we cannot definitively determine whether they increase or decrease in direct proportion to changes in the sample size (see Table 1).

2. The optimum bandwidth for bivariate and univariate varies significantly from one another (see Table 1).

Algorithm 1: Calculate the smoothing parameters for bivariate circular data using the Jones-Pewsey kernel

Data: Bivariate circular data $(\varphi_1, \psi_1)\,,\ (\varphi_2, \psi_2)\,,\ \ldots,\ (\varphi_n, \psi_n)$
Result: Optimal bandwidth for the data
Input: $n, \mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_0$

Generate the data from $BvM\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\ \begin{pmatrix}\kappa_1 & 0\\ 0 & \kappa_2\end{pmatrix}\right)$;

Compute;
$M_1 < -outer(phi, phi,\ " - ")$;
$M_2 << -\ outer(psi, psi,\ " - ")$;
begin
  $f_1 < function(\mathcal{B}_1)$;
  if $(\mathcal{B}_1 > 0)$ **then**
    $M_x < -\ djonespewsey\ (M_1, \mu\ = circular(0), kappa = \mathcal{B}_1,\ psi = -0.6)$;
    for $(i\ in\ 1 : n)$ **do**
      $\{M_x[i, i] = 0\}$;
      $f_{hat} < -\ apply(M_1, 1,\ sum)$;
      $f_{hat1} < -f_{hat}/(n-1)$;
      $\log .f_{hat} < -\log(f_{hat1})$;
      $sum.\log < -sum(\log .f_{hat1})$;
      $result < -(-sum.\log)$;
    end
  end
  return *(result)*;
  Apply Non-Linear Minimization (f_1, κ_0);
end

begin
$f_2 <- function(\mathcal{B}_2)$;
if $(\mathcal{B}_2 > 0)$ **then**
$M_y <- djonespewsey\ (M_2, \mu = circular(0), kappa = \mathcal{B}_2,\ psi = -0.6)$;
for $(i\ in\ 1 : n)$ **do**
$\{M_y[i, i] = 0\}$;
$f_{hat} <- apply(M_y, 1, sum)$;
$f_{hat2} <- f_{hat}/(n-1)$;
$\log .f_{hat} <- \log(f_{hat2})$;
$sum.\log <- sum\ (\log .f_{hat2})$;
$result <- (-sum.\log)$;
end
end
return *(result)*;
Apply Non-Linear Minimization (f_2, κ_0);
end
begin
$f_3 <- function(\mathcal{B})$;
if $(\mathcal{B}[1] > 0 \& \mathcal{B}[2] > 0)$ **then**
$\mathcal{B}_1 <- \mathcal{B}[1]$;
$\mathcal{B}_1 <- \mathcal{B}[2]$;
$M_x <- djonespewsey(M_1, \mu = circular(0), kappa = \mathcal{B}_1, psi = -0.6)$;
$M_y <- djonespewsey(M_2, \mu = circular(0), kappa = \mathcal{B}_2, psi = -0.6)$;
$M <- (M_x * M_y)$;
for $(i\ in\ 1 : n)$ **do**
$M[i, i] = 0$
$g_{hat} <- apply(M, 1, sum)$;
$g_{hat1} <- g_{hat}/(n-1)$;
$\log .g_{hat} << -\log(g_{hat1})$;
$sum.\log <- sum\ (\log .g_{hat})$;
$result\ <- (-sum.\log)$;
end
end
end
$\mathcal{B}_x <- outx\$estimate$; $\mathcal{B}_y <- outy\$estimate$;
$c1 <- \mathcal{B}_x/(n^(2/5))$; $c2 <- \mathcal{B}_y/(n^(2/5))$;
$\mathcal{B}\mathcal{B}_x <- c1 * n^(2/6)$; $\mathcal{B}\mathcal{B}_y <- c2 * n^(2/6)$;
Apply Non-Linear Minimization $(f_3, (\mathcal{B}\mathcal{B}_x, \mathcal{B}\mathcal{B}_y))$

Table 1: The smoothing parameter for the two models mentioned above with varying sample sizes and the kernel function Jones-Pewsey distribution.

Model 1			
Sample size	Bandwidth for		Bandwidth for
	φ	ψ	(φ, ψ)
10	2.0323	3.5132	1.2841 3.5790
20	2.6833	0.3204	2.7184 0.4839
30	2.0829	0.6255	2.0648 0.3472
40	2.4825	1.6165	2.3982 1.4168
50	2.6509	1.6888	2.3418 1.3055
60	2.7337	2.2605	2.6495 2.0725
70	2.9528	2.8189	3.0802 1.8466
80	2.8711	2.7875	2.6266 1.4156
90	2.8467	1.9644	2.5298 1.3876
100	2.6722	1.7636	2.5643 1.5707
200	87.8577	2.683	61.7154 0.3032
300	43.3811	3.0932	29.6633 0.5354
400	38.2924	2.6937	25.7071 0.6306
500	42.9468	2.7853	28.3661 0.3635
1000	76.7275	26.8261	48.3944 83.3726
Model 2			
10	1.7065	3.2129	0.9808 2.8722
20	2.4862	0.4328	2.4561 0.3652
30	1.8513	1.0763	1.8629 0.8330
40	2.1664	0.0614	1.7694 0.0000
50	2.3649	1.0571	2.2720 0.5429
60	2.6155	1.2214	2.6682 0.9631
70	2.8177	1.9813	2.8793 0.5966
80	2.8016	2.7575	2.3035 2.5219
90	2.8861	2.2504	3.3284 0.7700
100	2.7002	2.4967	2.4627 1.2115
200	2.7627	1.5522	2.5503 1.0407
300	2.5901	1.7478	2.4290 1.4517
400	41.4635	2.196	27.8326 0.1094
500	34.4331	2.3758	26.2370 0.1963
1000	29.0381	2.509	25.2695 0.1376

4 Conclusion

A methodology for determining the optimal bandwidth for bivariate angular data points has been presented in this paper. The optimization technique (non-linear minimization) with kernel density estimation with the Jones-Pewsey distribution as the kernel function are used to construct this approach. Moreover, a simulation study has been presented to assess the proposed methodology across various models and sample sizes utilizing R software by generating data from a von-Mises circular distribution. According to the simulations results, our results are as follows: The joint pdf appears differently when the parameters are optimal compared to when they are random, suggesting that the joint pdf is parameter-dependent for bivariate directional data. The free parameters (bandwidth) in the kernel function significantly affect pdf because these parameters control the behavior and flexibility of the kernel function.

References

[1] J. Cremers, I. Klugkist, One direction? A tutorial for circular data analysis using R with examples in cognitive psychology, Frontiers in psychology, **9,** (2011), 396689, 2018.

[2] S. Zámečník, I. Horová, S. Katina, K. Hasilová, An adaptive method for bandwidth selection in circular kernel density estimation, Computational Statistics, **39,** no. 4, (2024), 1709–1728.

[3] C. C. Taylor, Automatic bandwidth selection for circular density estimation, Computational Statistics & Data Analysis, **52,** no. 7, (2008), 3493–3500.

[4] 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.

[5] M. Oliveira, R. M. Crujeiras, A. Rodríguez-Casal, A plug-in rule for bandwidth selection in circular density estimation, Computational Statistics & Data Analysis, **56,** no. 12, (2012), 3898–3908.

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

[7] S. F. H. Abushilah, Statistical approach for bivariate circular data with application to bioinformatics, MSc thesis, University of Kufa.

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

[9] K. L. Richards, Design Engineer's Reference Guide: Mathematics, Mechanics, and Thermodynamics, CRC Press, Florida, 2014.

[10] B. D. Ripley, Modern Applied Statistics with S, Springer, New York, 2002.