STATISTICS IN TRANSITION new series, September 2024
Vol. 24, No. 3, pp. 1–12, https://doi.org/10.59170/stattrans-2024–024
Received – 14.08.2023; accepted – 05.06.2024

A comprehensive exploration of complete cross-validation for circular data

Kamila Hasilová[1], Ivana Horová[2], David Vališ[3], Stanislav Zámečník[4]

Abstract

Kernel density estimation of circular data has recently received considerable attention for its ability to model and analyse distributions on unit circles and other periodic domains. Our aim is to contribute to the literature on data-driven bandwidth selectors in circular kernel density estimation. We propose a novel circular-specific method that is based on a cross-validation procedure with a von Mises density used as a kernel function. Using simulated data as well as real-world circular datasets, we evaluate and validate the proposed method and compare it with the existing methods.

Key words: circular data, kernel density estimation, von Mises density, cross-validation method.

1. Introduction

The analysis of data in terms of directions in a plane/space, or equivalently in terms of positions of points on a circle/sphere, is required in many content areas in the Earth sciences, astrophysics, social sciences, and other fields. This type of data – known as circular data – is generally defined by the main circular measuring devices – the compass and the clock. Wind directions, animal movements, biological and social rhythms, times of occurrence of an event are illustrative examples of such observations (Mardia and Jupp, 2000; Ley and Verdebout, 2017).

The treatment of these data can also be realised by means of kernel smoothing methods. Kernel smoothing belongs to a category of nonparametric curve estimation techniques. The aim of kernel smoothing is, i.a., to estimate the entire probability density function with as few assumptions about the density as possible (Wand and Jones, 1995).

The main issue here is bandwidth selection. Among the earliest fully automatic and data-driven bandwidth selection methods are those based on cross-validation ideas (see, e.g. Rudemo, 1982; Bowman, 1984; Silverman, 1986; Scott, 1992). The main idea of these methods is a natural approximation to the mean integrated square error.

[1]Department of Quantitative Methods, Faculty of Military Leadership, University of Defence, Brno, Czech Republic. E-mail: kamila.hasilova@unob.cz. ORCID: https://orcid.org/0000-0003-1540-3489.
[2]Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, Czech Republic. E-mail: horova@math.muni.cz.
[3]Department of Combat and Special Vehicles, Faculty of Military Technology, University of Defence, Brno, Czech Republic. E-mail: david.valis@unob.cz.
[4]Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, Czech Republic. E-mail: zamecnik.stanislav@gmail.com.

Jones and Kappenman (1991) gave an overview of bandwidth selection methods: least square cross-validation, biased cross-validation, plug-in, and presmoothed cross-validation for linear data. Further, they proposed a new type of cross-validation. The new method is called complete cross-validation. This procedure estimates the entire mean integrated square error as opposed to least square cross-validation estimation, which omits the part depending on the unknown density.

The main point of this paper is to apply a modification of the above mentioned method to the circular data. It is not a straightforward task because circular data are fundamentally different from linear data, there is no true zero, any classification of low or high values is arbitrary. In addition, the periodic nature of the circular data complicates their analysis, as standard methods for linear data in Euclidean space are inappropriate for circular data analysis.

2. Circular kernel density estimation

Let $\Theta_1,\ldots,\Theta_n \in [0,2\pi)$ be a random sample of angles drawn from a distribution with a density function f. The kernel estimate of this density at a point (angle) θ is defined as

$$\widehat{f}(\theta,\nu) = \frac{1}{n}\sum_{i=1}^{n} K_\nu(\theta-\Theta_i), \tag{1}$$

where $K_\nu(\cdot)$ is a circular kernel function and $\nu > 0$ is a concentration parameter playing the role of a smoothing parameter. For parameter ν it is required that $\nu\to\infty$, $\sqrt{\nu}n^{-1}\to 0$ for $n\to\infty$ (Oliveira et al., 2012; Tsuruta and Sagae, 2017).

As a circular kernel, the von Mises density function can be considered. The von Mises distribution, $vM(\mu,\kappa)$, is a symmetric unimodal distribution with a mean direction, $\mu \in [0,2\pi)$, and a concentration parameter, $\kappa > 0$. Its density function takes the following form

$$g(\theta;\mu,\kappa) = \left[2\pi I_0(\kappa)\right]^{-1}\exp(\kappa\cos(\theta-\mu)),$$

where $I_0(\kappa)$ is the modified Bessel function of order zero (Jammalamadaka and SenGupta, 2001).

A critical issue in the application of this estimator in practice is the selection of the smoothing parameter which controls the smoothness of the estimate. Small values of the smoothing parameter imply less concentrated data and provide oversmoothed estimates. On the other hand, large values result in undersmoothed estimates, which may reveal local structure in the data.

Data-driven procedures for selecting ν are usually based on error performance measures. One of these is the bandwidth minimising the mean integrated square error (MISE) of $\widehat{f}$, which is defined as

$$\mathrm{MISE}(\nu) = E\int_0^{2\pi}\left[\widehat{f}(\theta,\nu)-f(\theta)\right]^2 \mathrm{d}\theta.$$

Denoting $R(w,z) = E\int_0^{2\pi} w\cdot z\,\mathrm{d}\theta$ and $R(z)=R(z,z)$, we can write

$$\mathrm{MISE}(\nu) = R(\widehat{f}) - 2R(\widehat{f},f) + R(f). \tag{2}$$

3. Bandwidth selection

The above expression does not give a practical procedure for choosing the optimal smoothing parameter, because some of the MISE terms depend on the unknown density f. However, one can estimate these f-dependent quantities by those that include $\widehat{f}$ and then select the optimal bandwidth (Hall and Marron, 1987).

In the following computations, we restrict ourselves to using the von Mises density as the kernel function, i.e. $K_\nu(\theta) = \left[2\pi I_0(\nu)\right]^{-1} \cdot \exp(\nu\cos\theta)$, since it plays a similar role in the circular case as the Gaussian density does in the linear case.

Let us denote the ratio of the Bessel functions as $A_k(\nu) = I_k(\nu)/I_0(\nu)$ and the difference between the sample values as $\Theta_{ij} = \Theta_i - \Theta_j$. This notation is useful for the derivations and calculations in the following sections.

3.1. Least square cross-validation

The least square cross-validation selector aims to minimise (2) with the term $R(f)$ dropped. It is defined as the minimum of the function

$$\mathrm{LSCV}(\nu) = R(\widehat{f}) - \frac{2}{n}\sum_{i=1}^{n}\widehat{f}_{-i}(\Theta_i,\nu),$$

where $\widehat{f}_{-i}$ denotes the estimate (1) with the i-th observation omitted

$$\widehat{f}_{-i}(\Theta_i,\nu) = \frac{1}{n-1}\sum_{\substack{i=1\\ j\neq i}}^{n} K_\nu(\Theta_{ij}), \quad i = 1,\ldots,n.$$

Rewriting the formulas using a kernel function, the LSCV objective function takes the form (where $*$ denotes the convolution)

$$\mathrm{LSCV}(\nu) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(K_\nu * K_\nu)(\Theta_{ij}) - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K_\nu(\Theta_{ij}). \tag{3}$$

It is possible to prove that (3) is an unbiased estimator of the quantity $\mathrm{MISE}(\nu) - R(f)$ (Hall et al., 1987).

3.2. Complete cross-validation

Jones and Kappenman (1991) investigated a class of data-driven bandwidth selection procedures for linear kernel density estimation. They proposed a new method (from the family of cross-validation methods) called complete cross-validation (CCV). Their proposal estimates the entire MISE function, as opposed to the LSCV method, which targets the difference $\mathrm{MISE}(\nu) - R(f)$.

Here, we present an adaptation of the CCV method to the circular case. Let us defined functionals T_m in a similar manner as those specified by Jones and Kappenman (1991):

$$T_m(\nu) = (-1)^m [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} K_\nu^{(2m)}(\Theta_{ij}),$$

where $K_\nu^{(2m)}$ is the $(2m)$-th derivative of K_ν. Then, we define the CCV function as

$$\mathrm{CCV}(\nu) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (K_\nu * K_\nu)(\Theta_{ij}) - T_0(\nu) + \frac{A_1(\nu)}{2\nu} T_1(\nu) + \frac{2A_1^2(\nu) - A_2(\nu)}{8\nu^2} T_2(\nu).$$

Lemma. $E[\mathrm{CCV}(\nu)] = \mathrm{MISE}(\nu) + o(\nu^{-2})$.

Proof. See the Appendix. □

4. Small sample behaviour

We carried out a small simulation study to compare the proposed CCV method and known LSCV method with eight simulation scenarios, see Figure 1 and Table 1.

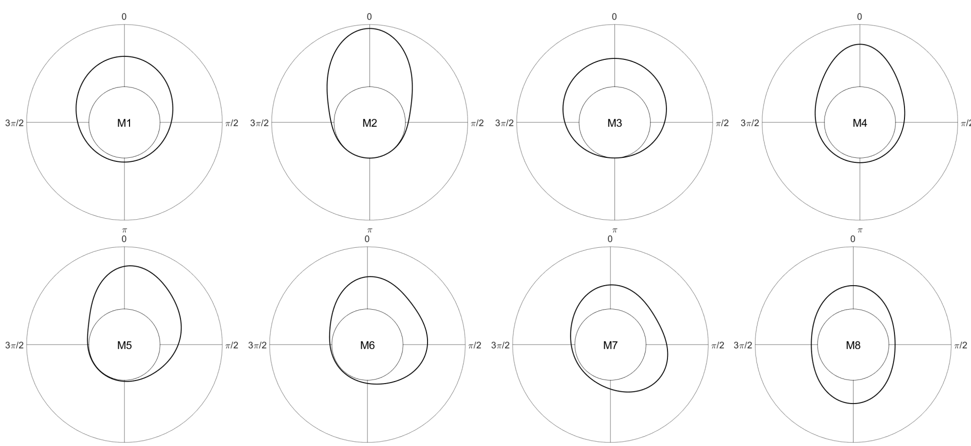

Figure 1: Graphical representation of models M1–M8

Table 1: Probability distributions and their parameters: von Mises (vM), cardioid (C), wrapped Cauchy (WC) and mixtures of two von Mises distributions

model	distribution	model	distribution
M1	$vM(0;1)$	M5	$0.5vM(0;4) + 0.5vM(\pi/3;2)$
M2	$vM(0;2.5)$	M6	$0.5vM(0;3) + 0.5vM(\pi/2;2)$
M3	$C(0;0.5)$	M7	$0.5vM(0;2) + 0.5vM(2\pi/3;2)$
M4	$WC(0;0.5)$	M8	$0.5vM(0;2) + 0.5vM(\pi;2)$

Table 2: Mean and standard deviation of ISE for all the simulation scenarios

model	method	$n = 50$	$n = 100$	$n = 200$
M1	LSCV	0.0150 (0.0226)	0.0083 (0.0096)	0.0046 (0.0046)
	CCV	0.0302 (0.0081)	0.0279 (0.0058)	0.0262 (0.0059)
M2	LSCV	0.0233 (0.0231)	0.0143 (0.0141)	0.0087 (0.0077)
	CCV	0.1197 (0.0670)	0.0904 (0.0770)	0.0525 (0.0717)
M3	LSCV	0.0135 (0.0215)	0.0076 (0.0093)	0.0044 (0.0051)
	CCV	0.0319 (0.0079)	0.0298 (0.0057)	0.0281 (0.0061)
M4	LSCV	0.0215 (0.0243)	0.0127 (0.0122)	0.0069 (0.0054)
	CCV	0.0545 (0.0165)	0.0491 (0.0199)	0.0405 (0.0251)
M5	LSCV	0.0218 (0.0272)	0.0125 (0.0132)	0.0073 (0.0071)
	CCV	0.0726 (0.0207)	0.0669 (0.0272)	0.0648 (0.0289)
M6	LSCV	0.0182 (0.0241)	0.0101 (0.0102)	0.0062 (0.0055)
	CCV	0.0415 (0.0100)	0.0388 (0.0108)	0.0363 (0.0123)
M7	LSCV	0.0173 (0.0201)	0.0093 (0.0082)	0.0057 (0.0045)
	CCV	0.0281 (0.0077)	0.0253 (0.0071)	0.0220 (0.0084)
M8	LSCV	0.0208 (0.0194)	0.0109 (0.0089)	0.0063 (0.0043)
	CCV	0.0212 (0.0097)	0.0143 (0.0105)	0.0074 (0.0075)

Taking f to be a von Mises distribution (M1, M2), cardioid distribution (M3), wrapped Cauchy distribution (M4), and mixture of two von Mises distributions (M5–M8), we generated 200 random samples of size $n = 50$, $n = 100$, and $n = 200$.

The performance and comparison of the presented circular density estimators, i.e. LSCV and CCV, is assessed by the integrated square error, $\mathrm{ISE} = \int(\widehat{f} - f)^2\,\mathrm{d}\theta$, which was calculated numerically using the trapezoidal rule with 500 points. The mean values of ISE (with standard deviations) are summarised in Table 2.

Even though the CCV method has better theoretical properties, the results show that when applied to simulated data, the ISE error is greater than for the LSCV method. This is due to the fact that the CCV function comes from the family of cross-validation functions based on the estimate of MISE. Cross-validation methods tend to underestimate the resulting function in the linear case, which, given the nature of the smoothing parameter, leads to an oversmoothed estimate in the circular case.

Sometimes, cross-validation objective functions have more than one local minimum (Wand and Jones, 1995), see Figure 2, where we denote these two possible outcomes as Type I (with one minimum) and Type II (with two minima). For Type II objective functions, ν_{CCV} should be taken to be the 'largest' local minimiser of CCV.

In Table 3, we summarise the percentage of the Type I CCV objective functions in the simulation scenarios. Type II is more common for data sets larger in size and more concentrated (see models M2, M4, and M8).

Comparing the bandwidth parameters of the two methods, we can see that the ratio $\nu_{\mathrm{CCV}}/\nu_{\mathrm{LSCV}}$ ranges on average from 0.3 to 0.75. Although this ratio appears to be quite large and may imply that the LSCV estimate is undersmoothed (or similarly, the CCV estimate is oversmoothed), we can obtain decent results, as can be seen in Figure 3, which shows selected random samples of size $n = 200$ from the model M5 and their kernel density estimates with the bandwidth chosen by both methods. These methods ($\nu_{\mathrm{LSCV}} = 27.81$, $\nu_{\mathrm{CCV}} = 13.34$) provide similar results for sample (a). For sample (b), the LSCV method

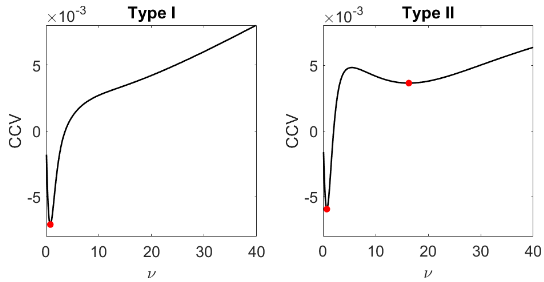

Figure 2: Two types of CCV objective functions

Table 3: Percentage of CCV function type I for the simulation models

model	$n = 50$	$n = 100$	$n = 200$
M1	93.5	93	91.5
M2	69.5	52.5	29.5
M3	92	93.5	91.5
M4	85.5	80	67
M5	82.5	80	79.5
M6	87	87.5	85
M7	89	86.5	81
M8	93.5	80.5	68

gives a very good estimate ($\nu_{\text{LSCV}} = 9.52$), but the CCV method gives an oversmoothed estimate ($\nu_{\text{CCV}} = 0.80$). However, the LSCV is not always an optimal method, as we can see in the graph of sample (c), which gives an undersmoothed LSCV estimate ($\nu_{\text{LSCV}} = 107.47$) and a very good CCV estimate ($\nu_{\text{CCV}} = 17.37$).

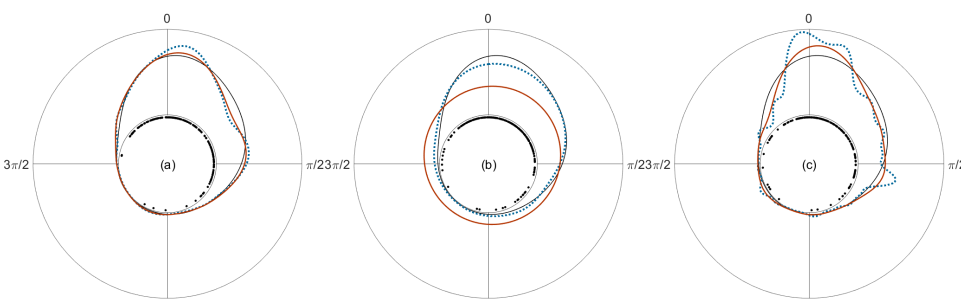

Figure 3: Selected samples from the model M5 and their estimates LSCV (blue, dotted), CCV (orange, full) with the true density (black, thin)

Admitting that there is no universally applicable method for circular kernel density estimation, we can see from the simulations that the CCV method is an acceptable alternative to the LSCV method. Looking at the graphs in Figure 3, we can consider applying the so-called averaging technique (Baszczyńska, 2017) to obtain the estimate that combines the local and global insight into the structure of the data.

5. Real data applications

The proposed method is applied to a few real data sets, each of them being similar to one of the models of the simulation study.

Example 1. Hisada (1972) studied dragonflies and their behaviour. The data set, available in the R package NPCirc (Oliveira et al., 2014), consists of 214 orientations of dragonflies with respect to the solar azimuth (measured in degrees). Figure 4(a) shows the data – zero corresponds to the azimuth of the sun – and the resulting kernel estimates obtained with optimal smoothing parameters $\nu_{\mathrm{LSCV}} = 63.87$ and $\nu_{\mathrm{CCV}} = 53.44$. As we can see, the CCV method gives almost the same density estimate as the LSCV method. As expected from the simulation study, the two methods provide almost the same results for data with two directly opposite modes (model M8).

Example 2. The Czech Hydrometeorological Institute has kindly provided the wind direction data. The data set represents the wind direction from Brno-Tuřany airport, the data were recorded every 10 minutes, resulting in 144 measurements for one day. For the analysis, 7 November 2021 was chosen. The optimal smoothing parameters obtained by the cross-validation methods are $\nu_{\mathrm{LSCV}} = 50.51$ and $\nu_{\mathrm{CCV}} = 30.77$, respectively. The data set shows bimodality (like model M7) and the resulting density estimates are similar, see Figure 4(b).

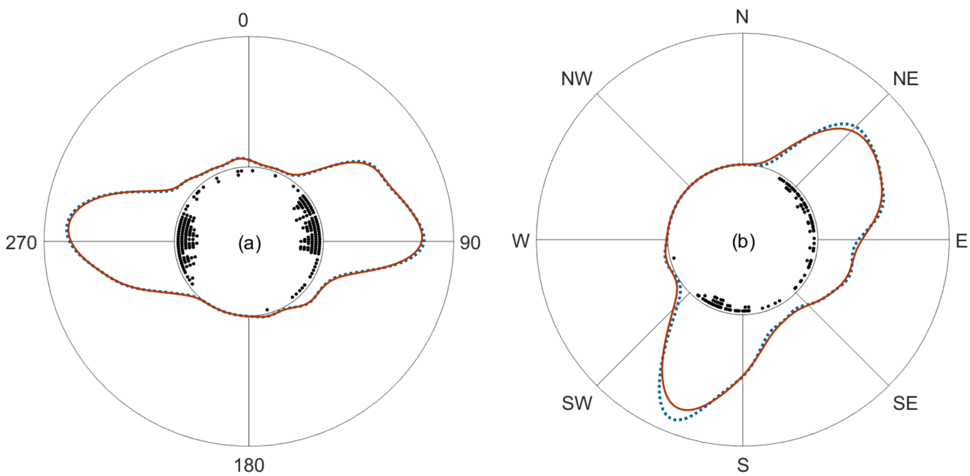

Figure 4: Kernel density estimates of the (a) dragonfly data and (b) wind data. Black points inside the circle represent the data, LSCV (dotted blue line) and CCV (full orange line) estimate

Example 3. In cooperation with the Lublin Municipal Transport Company, we collected data on bus subsystem failures over a period of more than seven years. The failures of each subsystem were recorded monthly, i.e. we only know the month of the specific failure, but not the exact date. The air conditioning (AC) subsystem ($n = 45$) was selected for investigation. The optimal smoothing parameters are $\nu_{\mathrm{LSCV}} = 2.37$ and $\nu_{\mathrm{CCV}} = 0.70$. Such small values leading to underestimated densities are due to the fact that these are aggregated data. Still, we can see that the failure occurrence is more likely to happen in early summer, i.e. in May, June, and July. The resulting estimated densities are displayed in Figure 5.

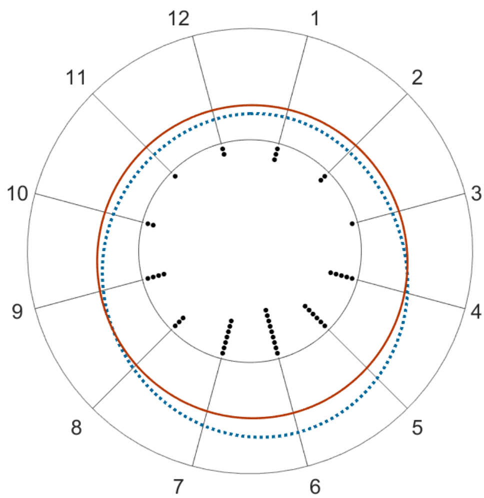

Figure 5: Kernel density estimate of the AC failure data (black points) with LSCV estimate (dotted blue) and CCV estimate (full orange)

6. Conclusion

We proposed a new complete cross-validation (CCV) method of bandwidth selection in circular density estimation. This method extends the estimate of the MISE and has better theoretical properties than the LSCV. From the presented results and outcomes we can conclude that the CCV is applicable to various data types with respective success rate.

From the data-driven method of bandwidth selection, we focus only on the cross-validation methods which target the MISE to have the consistent group of methods to compare. We have considered alternative approaches (see, e.g. Taylor, 2008; Oliveira et al., 2012; García-Portugués, 2013), but they target different type of error function – the asymptotic mean integrated square error. From the theoretical point of view also the augmented cross-validation (Tsuruta and Sagae, 2020), but this one cannot be used in real data application.

Although the CCV presents a theoretical concept, our aim is to make the method applicable to practical situations. The form and structure of the original datasets (corresponding to the true density) have to be taken into account. The size of the data set also plays an important role in producing results with some degree of validity. In datasets where data records are short, sparse or even missing, we cannot expect relevant estimates and results.

However, we are confident that the CCV method provides realistic and applicable results, not only in terms of key statistical results such as circular kernel density estimation, but also in terms of other measures that can be successfully applied in other areas such as reliability and safety engineering.

Acknowledgements

This paper has been prepared with the support of the Ministry of Defence of the Czech Republic, Partial Projects for Institutional Development LANDOPS and VAROPS, University of Defence, Brno, and Grant Agency of Masaryk University, project MUNI/A/1418/2022 "Mathematical and statistical modelling 4 (MaStaMo4)". We gratefully acknowledge the Czech Hydrometeorological Institute of Brno and the Lublin Municipal Transport Company for providing the wind direction and bus failure datasets, respectively.

References

Baszczyńska, A., (2017). One value of smoothing parameter vs interval of smoothing parameter values in kernel density estimation. *Acta Universitatis Lodziensis. Folia Oeconomica*, 6, pp. 73–86.

Bowman, A. W., (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, pp. 353–360.

García-Portugués, E., (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7, pp. 1655–1685.

Hall, P., Marron, J. S., (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, 6, pp. 109–115.

Hall, P., Watson, G. S., Cabrera, J., (1987). Kernel density estimation with spherical data. *Biometrika*, 74, pp. 751–762.

Hisada, M., (1972). Azimuth orientation of the dragonfly (*Sympetrum*). In Galler, S. R., et al. *Animal Orientation and Navigation*, pp. 511–522. Washington: U.S. Government Printing Office.

Jammalamadaka, S. R., SenGupta, A., (2001). *Topics in circular statistics*, Singapore: World Scientific.

Jones, M., Kappenman, R., (1991). On a class of kernel density estimate bandwidth selectors. *Scandinavian Journal of Statistics*, 19, pp. 337–349.

Ley, C., Verdebout, C., (2017). *Modern directional statistics*, Boca Raton: CRC Press.

Mardia, K. V., Jupp, P. E., (2000). *Directional statistics*, Chichester: Wiley.

Oliveira, M., Crujeiras, R. M., Rodríguez-Casal, A., (2012). A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data Analysis*, 56, pp. 3898–3908.

Oliveira, M., Crujeiras, R. M., Rodríguez-Casal, A., (2014). NPCirc: An R Package for Nonparametric Circular Methods. *Journal of Statistical Software*, 61, pp. 1–26.

Rudemo, M., (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, pp. 65–78.

Scott, D. W., (1992). *Multivariate density estimation: Theory, practice, and visualization*, New York: Wiley.

Silverman, B. W., (1986). *Density estimation for statistics and data analysis*, London: Chapman and Hall.

Taylor, C. C., (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*, 52, pp. 3493–3500.

Tsuruta, Y., Sagae, M., (2017). Higher order kernel density estimation on the circle. *Statistics and Probability Letters*, 131, pp. 46–50.

Tsuruta, Y., Sagae, M., (2020). Theoretical properties of bandwidth selectors for kernel density estimation on the circle. *Annals of the Institute of Statistical Mathematics*, 72, pp. 511–530.

Wand, M. P., Jones, M. C., (1995). *Kernel smoothing*, London: Chapman and Hall.

Appendix

Convolution of the von Mises kernel $K_\nu(\theta)$ with itself is (Jammalamadaka and SenGupta, 2001):

$$(K_\nu * K_\nu)(\theta) = \frac{I_0\left(\sqrt{2\nu^2(1+\cos\theta)}\right)}{2\pi I_0^2(\nu)}.$$

Functionals T: Derivatives of the von Mises kernel can be expressed using the recurrence relationship

$$K_\nu(\theta) = [2\pi I_0(\nu)]^{-1}\exp(\nu\cos\theta) \qquad P_1(\theta) = -\nu\cos\theta$$
$$K_\nu^{(n)}(\theta) = K_\nu(\theta)\cdot P_n(\theta) \qquad P_{n+1}(\theta) = -\nu\sin\theta\cdot P_n(\theta) + P_n'(\theta).$$

Then, the exact form of functionals T_0, T_1 and T_2 with the von Mises kernel is

$$T_0 = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K_\nu(\Theta_{ij}) = \frac{1}{2\pi I_0(\nu)n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\exp(\nu\cos\Theta_{ij}),$$

$$\begin{aligned}
T_1 &= -\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K_\nu^{(2)}(\Theta_{ij})\\
&= -\frac{1}{2\pi I_0(\nu)n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\exp(\nu\cos\Theta_{ij})\cdot\left(\nu^2\sin^2\Theta_{ij} - \nu\cos\Theta_{ij}\right),
\end{aligned}$$

$$\begin{aligned}
T_2 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K_\nu^{(4)}(\Theta_{ij})\\
&= \frac{1}{2\pi I_0(\nu)n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\exp(\nu\cos\Theta_{ij})\cdot\Big(\nu^4\sin^4\Theta_{ij} - 6\nu^3\sin^2\Theta_{ij}\cos\Theta_{ij}\\
&\qquad + 3\nu^2(\cos^2\Theta_{ij} - \sin^2\Theta_{ij}) - \nu^2\sin^2\Theta_{ij} + \nu\cos\Theta_{ij}\Big).
\end{aligned}$$

Proof of the Lemma: First, we show expectations of functionals $T_m(\nu)$.

$$\begin{aligned}
ET_0(\nu) &= E\left[K_\nu(\theta_1-\theta_2)\right] = R(f) - \frac{A_1(\nu)}{2\nu}R(f') + \frac{A_2(\nu)}{8\nu^2}R(f'') + o(\nu^{-2}),\\
ET_1(\nu) &= E\left[-K_\nu''(\theta_1-\theta_2)\right] = R(f') - \frac{A_1(\nu)}{2\nu}R(f'') + o(\nu^{-1}),\\
ET_2(\nu) &= E\left[K_\nu^{(4)}(\theta_1-\theta_2)\right] = R(f'') + o(1).
\end{aligned}$$

Thus

$$
\begin{aligned}
E[\mathrm{CCV}(\nu)] &= E[R(\widehat{f})] - ET_0(\nu) + \frac{A_1(\nu)}{2\nu} ET_1(\nu) + \frac{1}{8\nu^2}\big(2A_1^2(\nu) - A_2(\nu)\big) ET_2(\nu) + o(\nu^{-2}) \\
&= E[R(\widehat{f})] - R(f) + \frac{A_1(\nu)}{\nu} R(f') - \frac{A_2(\nu)}{4\nu^2} R(f'') + o(\nu^{-2}).
\end{aligned}
$$

On the other hand, the middle term of MISE can be expressed as follows:

$$
\begin{aligned}
E\int \widehat{f}(\theta,\nu) f(\theta)\,\mathrm{d}\theta &= \iint K_\nu(\theta-\alpha) f(\alpha) f(\theta)\,\mathrm{d}\alpha\,\mathrm{d}\theta \\
&= R(f) - \frac{A_1(\nu)}{2\nu} R(f') + \frac{A_2(\nu)}{8\nu^2} R(f'') + o(\nu^{-2})
\end{aligned}
$$

and the whole MISE reads as

$$
\begin{aligned}
\mathrm{MISE} &= E[R(\widehat{f})] - 2E\int \widehat{f}(\theta,\nu) f(\theta)\,\mathrm{d}\theta + R(f) \\
&= E[R(\widehat{f})] - 2R(f) + 2\frac{A_1(\nu)}{2\nu} R(f') - 2\frac{A_2(\nu)}{8\nu^2} R(f'') + R(f) + o(\nu^{-2}) \\
&= E[R(\widehat{f})] - R(f) + \frac{A_1(\nu)}{\nu} R(f') - \frac{A_2(\nu)}{4\nu^2} R(f'') + o(\nu^{-2}).
\end{aligned}
$$