Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 52 (2008) 3493–3500
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Automatic bandwidth selection for circular density estimation

Charles C. Taylor*

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

Received 29 May 2007; received in revised form 6 November 2007; accepted 9 November 2007
Available online 17 November 2007

Abstract

Given angular data $\theta_1, \ldots, \theta_n \in [0, 2\pi)$ a common objective is to estimate the density. In case that a kernel estimator is used, bandwidth selection is crucial to the performance. A "plug-in rule" for the bandwidth, which is based on the concentration of a reference density, namely, the von Mises distribution is obtained. It is seen that this is equivalent to the usual Euclidean plug-in rule in the case where the concentration becomes large. In case that the concentration parameter is unknown, alternative methods are explored which are intended to be robust to departures from the reference density. Simulations indicate that "wrapped estimators" can perform well in this context. The methods are applied to a real bivariate dataset concerning protein structure.

1. Introduction

Given a random sample of angles $\theta_1, \ldots, \theta_n \in [0, 2\pi)$ from some unknown density $f(\theta)$ a natural component of exploratory data analysis is to estimate the function $f(\cdot)$. When a parametric form is assumed, this may be achieved by maximum likelihood, or moment-based estimation. A nonparametric estimator may be naively written as

$$\hat{f}(\theta; h) = \frac{1}{n}\sum_{i=1}^{n} K_h(\theta - \theta_i), \tag{1}$$

where $K_h(\theta) = K(\theta/h)/h$ is a kernel function, usually a symmetric probability density, and h is a smoothing parameter. This kernel estimator was first proposed by Fisher (1989) for data lying on the circle, in which he adapted the linear data methods of Silverman (1986) and used a quartic kernel function $K(\theta) = 0.9375(1-\theta^2)^2$. However, when using data on the circle, we cannot use distance in Euclidean space, so all differences $\theta - \theta_i$ should be replaced by considering the angle between two vectors:

$$d_i(\theta) = \|\theta - \theta_i\| = \min(|\theta - \theta_i|, 2\pi - |\theta - \theta_i|). \tag{2}$$

This may also be written as $d_i = \cos^{-1}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}_i)$, where $\boldsymbol{x}^{\mathrm{T}} = (\cos\theta, \sin\theta)$ is a unit vector. A more natural choice for the kernel function is therefore one of the commonly used circular probability densities, such as the wrapped normal distribution, or the von Mises distribution. This leads to an alternative representation for the kernel density estimate (Jammalamadaka and SenGupta, 2001, page 282):

$$\hat{f}(\boldsymbol{x}; h) = \frac{1}{n}\sum_{i=1}^{n} K_h(1 - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}_i). \tag{3}$$

* Tel.: +44 (0) 113 343 5168; fax: +44 (0) 113 343 5090.
E-mail address: c.c.taylor@leeds.ac.uk.

0167-9473/$ - see front matter
doi:10.1016/j.csda.2007.11.003

In studying properties of kernel density estimates in Euclidean space, it is common to take Taylor series approximations to give an asymptotic form for the bias and variance. These can then be combined to give an asymptotically optimal choice for the smoothing parameter; see, for example, Silverman (1986). For data lying on the q-dimensional sphere ($q \geq 2$), Hall et al. (1987) described the asymptotic bias and variance of two classes of kernel estimators. This was done by the use of directional derivatives, thus making the results a close analogue of the Taylor series methods used for data in Euclidean space.

One of the difficulties in nonparametric density estimation is to make good choices of the smoothing parameter h; see Jones et al. (1996) for an excellent survey of methods. In the Euclidean setting, Silverman (1986) and Jones et al. (1996) gave formulae which depend on derivatives of the unknown density f. When the data lie in Euclidean space, there are many approaches to this problem, a simple example of which is based on a "Normal-scale rule" or a "rule of thumb". When the kernel function is taken as the Gaussian density, this leads to a plug-in selector $h = 1.06\hat{\sigma}n^{-1/5}$ (Silverman, 1986). The goal of this paper is to obtain an equivalent plug-in rule for density estimation on the circle.

Specifically, we consider the estimator in which the kernel function is the von Mises density, which gives

$$\hat{f}(\theta;\nu) = \frac{1}{n(2\pi)I_0(\nu)} \sum_{i=1}^{n} \exp\{\nu\cos(\theta-\theta_i)\}, \tag{4}$$

where $I_r(\nu)$ is the modified Bessel function of order r, and the concentration parameter ν has now taken the role of the (inverse of the) smoothing parameter h. A common approach to obtain the smoothing parameter is by considering derivatives of the unknown density and then substituting a "reference" density in order to obtain a plug-in rule; the results of Klemelä (2000) could probably be implemented here. However, we instead follow the approach of Marron and Wand (1992) who obtained the form of the exact mean integrated squared error for densities which can be expressed as a mixture of normal densities.

In Section 2 we write the exact expectation and variance of the estimator (4) under the assumption that the data follow a von Mises distribution. This then leads to an expression for the asymptotic bias and variance, which can be integrated to give AMISE as a function of the concentration parameter of the data (κ), the smoothing parameter (ν) and the sample size (n). Finally, this can be solved to give a simple plug-in rule for ν dependent only on κ and n. Section 3 discusses robust estimation of κ, suited for the plug-in rule, which may be used in case that the underlying density is not von Mises. Section 4 gives some simulation results, and Section 5 gives a real example using 2-dimensional data from a bioinformatics dataset. We conclude with a discussion.

2. Asymptotic mean integrated squared error

We suppose that $f(\cdot)$ is von Mises (written in general as $\mathrm{vM}(\mu,\kappa)$), with concentration parameter κ and – without loss of generality – mean direction $\mu = 0$. Then the first two moments of (4) are given by

$$\begin{aligned}\mathsf{E}\{\hat{f}(\theta;\nu)\} &= \frac{1}{(2\pi)^2 I_0(\kappa)I_0(\nu)} \int_0^{2\pi} \exp\{\nu\cos(\theta-\phi)+\kappa\cos(\phi)\}\mathrm{d}\phi \\ &= \frac{I_0\{(\kappa^2+\nu^2+2\nu\kappa\cos\theta)^{1/2}\}}{(2\pi)I_0(\kappa)I_0(\nu)},\end{aligned}$$

(Jammalamadaka and SenGupta, 2001, p. 40) and

$$\begin{aligned}\mathsf{var}\{\hat{f}(\theta;\nu)\} &= \frac{1}{n(2\pi)^2 I_0(\nu)^2}\,\mathsf{var}[\exp\{\nu\cos(\theta-\Theta)\}] \\ &= \frac{1}{n(2\pi)^2 I_0(\nu)^2 I_0(\kappa)}\left[I_0\{(4\nu^2+\kappa^2+4\kappa\nu\cos\theta)^{1/2}\} - \frac{I_0\{(\nu^2+\kappa^2+2\kappa\nu\cos\theta)^{1/2}\}^2}{I_0(\kappa)}\right].\end{aligned}$$

Note that, when $\nu = 0$ we have $\mathsf{E}\{\hat{f}(\theta;0)\} = 1/(2\pi)$ which does not depend on θ and, in the limit, the estimator is unbiased, *i.e.*

$$\lim_{\nu\to\infty} \mathsf{E}\{\hat{f}(\theta;\nu)\} = f(\theta).$$

These equations may be used to write down an expression for the exact mean squared error. However, integrating the resulting expression to obtain the exact MISE seems hard to do analytically, so we now derive asymptotic expressions for the above.

As the smoothing parameter $\nu \to \infty$ the asymptotic bias is

$$\{2\pi I_0(\kappa)\}^{-1}\left(\exp\left[\nu\left\{\left(1+\frac{\kappa^2}{\nu^2}+2\frac{\kappa}{\nu}\cos\theta\right)^{1/2}-1\right\}\right]-\exp\{\kappa\cos\theta\}\right)+O\left(\nu^{-2}\right).$$

Expanding the square root in a Taylor series, then expanding the exponential function in a Taylor series give a simpler form of the asymptotic bias as

$$\{4\pi I_0(\kappa)\nu\}^{-1}\kappa^2\sin^2\theta\exp(\kappa\cos\theta)+O\left(\nu^{-2}\right). \tag{5}$$

Similarly, for large n, and as $\nu \to \infty$ the variance has asymptotic form

$$\{4n\pi^{3/2}I_0(\kappa)\}^{-1}\nu^{1/2}\exp\left[2\nu\left\{\left(1+\frac{\kappa^2}{4\nu^2}+\frac{\kappa}{\nu}\cos\theta\right)^{1/2}-1\right\}\right]+o\left(\frac{\nu^{1/2}}{n}\right),$$

which is valid provided $n/\nu^{1/2} \to \infty$. Again, by expanding the square root, and then the exponential function, as a Taylor series, we obtain the simpler form of the asymptotic variance

$$\{4n\pi^{3/2}I_0(\kappa)\}^{-1}\nu^{1/2}\exp(\kappa\cos\theta)+o\left(\frac{\nu^{1/2}}{n}\right). \tag{6}$$

We now integrate the square of the asymptotic bias (5) and the asymptotic variance (6), to obtain

$$3\kappa^2 I_2(2\kappa)/\{32\pi\nu^2 I_0(\kappa)^2\}$$

and

$$\nu^{1/2}/\left(2n\pi^{1/2}\right)$$

respectively. Thus the asymptotic integrated mean squared error is of the form $a\nu^{-2}+b\nu^{1/2}$ which can be minimized by differentiating with respect to ν and equating to zero. This leads to a "von Mises-scale plug-in rule" for the smoothing parameter ν based on the estimated κ:

$$\nu=\left[3n\hat{\kappa}^2 I_2(2\hat{\kappa})\{4\pi^{1/2}I_0(\hat{\kappa})^2\}^{-1}\right]^{2/5}. \tag{7}$$

Note that this is of a similar asymptotic form as the normal-scale plug-in rule when we recall that ν is the concentration parameter, and so takes the role of $1/h^2$ in $h = 1.06\hat{\sigma}n^{-1/5}$. Moreover, if we consider the limit as $\kappa \to \infty$ then the von Mises distribution tends to a Normal distribution, with $\sigma = \kappa^{1/2}$. Hence, in the limit we have $h = \nu^{-1/2} = 1.06\kappa^{-1/2}n^{-1/5}$ which is exactly the same as the usual rule of thumb used for the Normal distribution. A simple method could be to estimate κ from the data, and use (7) to select the smoothing parameter for use in (4). Two obvious questions arise at this point: what happens if the data do not come from this reference density (von Mises); how good are all these Taylor series approximations in practice? The next two sections address these questions in turn.

3. Robust estimation of spread

When the data are unimodal, the above selection rule (7) is likely to work reasonably well. However, for bimodal data, the usual estimate of κ – either by maximum likelihood, or the method of moments – may be almost useless. In the most extreme case, an equal mixture of data tightly clustered around ϕ combined with a similar distribution of data clustered around $\phi+\pi$ will lead to an estimate of κ close to zero. When $\hat{\kappa}=0$ then (7) gives $\nu=0$ which will result in $\hat{f}(\theta) \equiv 1/(2\pi)$, and so such automatic methods may lead to very misleading density estimates. Indeed, even in the regular case, the maximum likelihood estimator of κ is far from robust, as it has infinite standardized gross error sensitivity (Mardia and Jupp, 1999, p. 276).

Ronchetti (1992) derived the "most B-robust self-standardized estimator" of the concentration parameter as $\hat{\kappa}_{\mathrm{MR}} = \log 2/\mathrm{median}\{c_i\}$, where $c_i = 1 - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\mu}$, with $\boldsymbol{\mu}$ the unit vector with direction μ. (In our implementation, we estimated μ with the mean direction.) Alternative robust estimators are also given by Ronchetti (1992) and Ko (1992), but our intention in this paper is to focus on density estimation.

In the case of Euclidean data, an alternative rule of thumb proposed by Silverman (1986, p. 47) was to take $\hat{\sigma} = \min\{s, \mathrm{IQR}/1.349\}$, where s is the sample standard deviation, and IQR is the inter-quartile range. This will work better for bimodal data, and give similar results when the data are normal. This proposal was obtained by comparing the population inter-quartile range to the standard deviation. For circular data, if m is the (estimated) median then, for $0 < p < 0.5$ define $q_i(p) \in [0, \pi)$ such that

$$p = \int_{m-q_1(p)}^{m} f(\theta)\mathrm{d}\theta = \int_{m}^{m+q_2(p)} f(\theta)\mathrm{d}\theta$$

which can be solved for known $f(\cdot)$ and given p. In particular, for the reference (von Mises) distribution, without loss of generality we can set $m = 0$. The inter p-quantile range for the reference distribution is then given by $q_2(p) + q_1(p) = 2q_1(p)$. The sample circular median is defined (Mardia and Jupp, 1999, p. 17) as the value $\hat{m}$ such that half the data lies in $[\hat{m}, \hat{m} + \pi)$ and more data lies closer to $\hat{m}$ than to $\hat{m} + \pi$. Sample values of $q_i(p)$ can then be easily found from the data. The procedure is then as follows:

(1) Select $p \in (0, 1/2)$.
(2) Form a look-up table which defines $q_1(p)$ as a function of κ for the reference distribution $\mathrm{vM}(0, \kappa)$.
(3) Find the sample median $\hat{m}$ and $\hat{q}_i(p)$, $i = 1, 2$ from the data.
(4) Obtain the estimated κ from the look-up table, using $\|\hat{m} + \hat{q}_2(p) - (\hat{m} - \hat{q}_1(p))\|$, where the distance used is as in (2).

An alternative approach is to note that, for the von Mises distribution, the maximum likelihood estimate of κ is obtained from the solution to

$$A_1(\kappa) = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i - \hat{\mu}),$$

where $A_k(\kappa) = I_k(\kappa)/I_0(\kappa)$ and $\hat{\mu} = \tan^{-1}(\sum \sin\theta_i, \sum \cos\theta_i)$. This follows from a more general identity using trigonometric moments which states that $\mathsf{E}\cos\{k(\theta - \mu)\} = I_k(\kappa)/I_0(\kappa)$. Thus, alternative estimates of κ (for a von Mises distribution) are given by solutions to

$$A_k(\kappa) = \frac{1}{n}\sum_{i=1}^{n}\cos(k\theta_i - \hat{\mu}_k), \tag{8}$$

where $\hat{\mu}_k = \tan^{-1}(\sum \sin k\theta_i, \sum \cos k\theta_i)$, for $k = 1, 2, \ldots$. In case that the data are von Mises, different values of k will lead to similar estimates of κ. In simulations (not shown), we have observed that $\hat{\kappa}_k$ is an increasing function of k, with the bias decreasing, and the variance increasing as κ increases. However, in the case of multimodal data, then rather different estimates will ensue. Hence, a possible procedure is to estimate κ using $k = 1, \ldots, K$ in (8) giving, say, $\hat{\kappa}_k$ and then taking $\hat{\kappa} = \max\{\hat{\kappa}_k, k = 1, \ldots, K\}$ for use in (7).

4. Simulations

For the standard von Mises distribution, we can compare the average integrated squared error $\overline{\mathrm{ISE}}$ with the approximate MISE given in Section 2, when κ is known. The results, for 500 simulations, and $n = 50$ and $n = 500$ are shown in Fig. 1. The approximation looks quite good, improving with n.

We now explore the effectiveness of the plug-in rule, when the data are taken from a mixture of $M(\geq 1)$ von Mises distributions. Specifically, we simulate $\theta_1, \ldots, \theta_n \sim f(\theta)$, where the distribution is given by

$$f(\theta) = \frac{1}{2\pi}\sum_{j=1}^{M} p_j \frac{\exp\{\kappa_j \cos(\theta - \mu_j)\}}{I_0(\kappa_j)}, \quad i = 1, \ldots, n \text{ with } \sum_{j=1}^{M} p_j = 1 \tag{9}$$

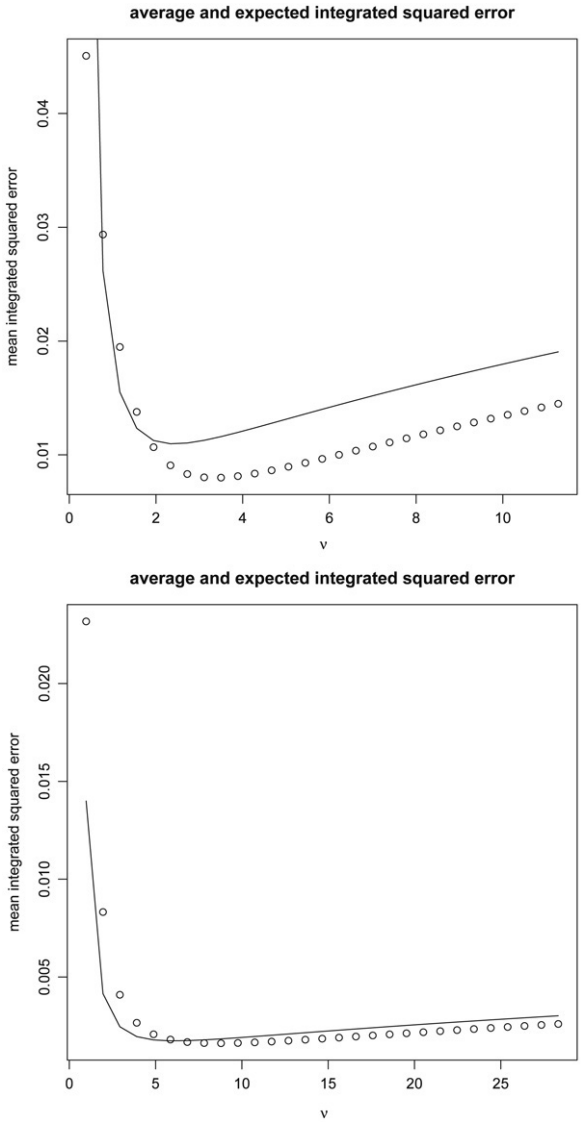


Fig. 1. Average integrated squared error – $\overline{\text{ISE}}$ – (points) and MISE (lines) for 500 simulations of size $n = 50$ (top panel) and $n = 500$ (bottom panel) from a von Mises distribution with $\kappa = 1$.

and we evaluate $\overline{\text{ISE}}(\nu) = \int (\hat{f}(\theta; \nu) - f(\theta))^2 \mathrm{d}\theta$ over $N = 500$ datasets (using a grid of 500 points to evaluate the integrals numerically). For each distribution, we note the value of ν which minimizes $\overline{\text{ISE}}(\nu)$, say ν_0, as well as $\overline{\text{ISE}}(\nu_0)$. We give $\overline{\text{ISE}}(\nu)$ when ν is obtained for each dataset from the plug-in rule (7) and $\hat{\kappa}$ is estimated by one of the methods described in Section 3. In addition, we give results when cross validation is used to select the bandwidth. Here, we select ν to maximize the likelihood cross-validation function $\text{LCV}(\nu) = \prod_i \hat{f}_{-i}(\theta_i; \nu)$, where

$$\hat{f}_{-i}(\theta; \nu) = \frac{1}{(n-1)(2\pi) I_0(\nu)} \sum_{j \neq i}^{n} \exp\{\nu \cos(\theta - \theta_j)\}$$

is the leave-one-out estimator. (We have also tried least-squares cross validation to select the smoothing parameter. The results of this were very similar to, but not quite as good as using likelihood cross validation, and so are not shown.) Let ν_{CV} denote the value of ν which maximizes $\text{LCV}(\nu)$. Denote by ν_K when ν is estimated with $\hat{\kappa} = \max\{\hat{\kappa}_k, k = 1, \ldots, K\}$ and $\hat{\kappa}_k$ is the solution to (8). Denote by ν_p the value of ν when κ is estimated using the inter p-quantile range. We also include results for Fisher's (1989) adaptation of the quartic kernel, in which his

Table 1
Average integrated squared error results for various bandwidth selection rules

Parameters	(1)		(0.1)		$\left(4, \frac{1}{2}, \pi, 4\right)$		$\left(2, \frac{1}{4}, \frac{\pi}{\sqrt{3}}, 2\right)$		$\left(5, \frac{1}{5}, \frac{\pi}{2}, 5\right)$	
n	50	500	50	500	50	500	50	500	50	500
ν_0	3.32	8.59	0.23	0.65	10.29	29.52	4.59	13.25	14.37	40.49
$100\,\overline{\mathrm{ISE}}(\nu_0)$	0.85	0.16	0.07	0.05	1.91	0.34	1.01	0.21	1.93	0.37
ν_{CV}	47.2	19.7	891.1	151.8	13.9	5.0	33.6	14.54	22.0	16.1
h_{F}	13.9	3.8	569.1	98.8	385.8	2857.5	6.7	−1.8	5.6	8.6
h_{E}	3.5	−3.9	487.4	50.0	431.7	3129.1	9.9	−9.3	16.4	20.9
ν_1	32.8	15.2	61.3	30.3	465.1	3263.1	25.1	4.4	28.2	51.7
ν_2	29.1	13.4	788.3	37.4	11.0	4.6	16.1	4.7	21.8	37.8
ν_3	36.6	12.8	1744.0	218.3	11.1	4.6	32.8	16.3	9.2	3.0
ν_4	54.7	21.5	2643.1	410.6	19.1	6.5	74.2	31.7	15.8	3.9
ν_{MR}	16.1	12.1	317.6	59.2	272.4	733.6	33.5	31.7	48.5	71.7
$\nu_{0.08}$	149.4	31.0	1229.0	20.3	182.4	1425.5	147.3	37.0	73.3	20.8
$\nu_{0.24}$	53.4	20.9	258.3	28.9	278.2	1982.9	57.6	32.9	29.5	18.0
$\nu_{0.40}$	38.4	17.4	143.4	32.5	372.4	2668.2	50.5	35.5	28.2	22.8

The parameters of the distribution, given in the top row by (9), are $(\kappa_1, p_2, \mu_2, \kappa_2, \ldots, p_M, \mu_M, \kappa_M)$, with $\mu_1 = 0$ in each case. Numerical integration used on 500 grid values; averages taken over 500 datasets of size n. ν_0 is the smoothing parameter to minimize MISE, and $\overline{\mathrm{ISE}}(\nu_0)$ the corresponding minimum. In the lower part of the table we give the percentage increase, *i.e.* $(\overline{\mathrm{ISE}}(\nu_\bullet)/\overline{\mathrm{ISE}}(\nu_0) - 1)100\%$ for each method. Here $\nu_\bullet$ is selected by cross validation (ν_{CV}), by ν_K, $K = 1, 2, 3, 4$ in the case where the wrapped estimator is used, by ν_p in the case where the p-quantile range estimator is used, and by ν_{MR} in the case where Ronchetti's (1992) robust estimator is used for κ. Two "linear" kernels are also used: h_{F} denotes the performance for the quartic kernel and respective plug-in rule described by Fisher (1989); h_{E} uses an Epanechnikov kernel with smoothing parameter $h_{\mathrm{E}} = 2.345\hat{\kappa}^{-1/2}n^{-1/5}$.

smoothing parameter is given by $h_{\mathrm{F}} = \sqrt{7}\hat{\kappa}^{-1/2}n^{-1/5}$, and for a similar method using the Epanechnikov kernel with $h_{\mathrm{E}} = 2.345\hat{\kappa}^{-1/2}n^{-1/5}$. This plug-in rule for a von Mises density was obtained by using a large concentration approximation for the AMISE solution given by

$$h_{\mathrm{E}} = \left(\frac{120\pi I_0(\kappa)^2}{n\kappa^2(2I_0(2\kappa) + I_2(2\kappa))}\right)^{1/5} \approx \kappa^{-1/2}(40\sqrt{\pi}/n)^{1/5}.$$

The results are given in Table 1. Note that, for the standard von Mises distribution, if the known $\kappa = 1$ is used in (7), then the smoothing parameter is $\nu = 3.51$ for $n = 50$ and $\nu = 8.82$ for $n = 500$, whereas if $\kappa = 0.1$ then $\nu = 0.06$ for $n = 50$ and $\nu = 0.16$ for $n = 500$, which shows the accuracy of the asymptotic results for finite samples. Note that using the maximum likelihood estimator for κ with (7) leads to row ν_1 in this table.

In Table 1 we see that for the standard von Mises distributions, only ν_1 (using (7) with κ estimated by (8) with $k = 1$) gives reasonable answers for both small and moderate κ. The linear kernel estimators are very poorly behaved for large smoothing parameters ($h > \pi$), which occur when $\hat{\kappa}$ and/or n are small. An *ad hoc* solution is simply to rescale the density estimate so that it integrates to unity, but this was not done here. However, note that for moderate κ (=1), the linear kernels outperform the von Mises kernel estimator. We conjecture that this may be due to the fact that the von Mises kernel is less efficient, though it is not immediately obvious how to define efficiency for angular kernels.

For the mixtures of distributions, the "standard" plug-in rule ν_1 can do very poorly, with both ν_2 and ν_3 performing similarly, overall, to the cross-validation estimate, but at a cheaper computational cost. Interestingly, the plug-in bandwidths for the linear kernels can perform surprisingly well for some of the mixtures. Amongst the p-quantile range estimators, $\nu_{0.40}$ performed reasonably, except for one of the mixture distributions.

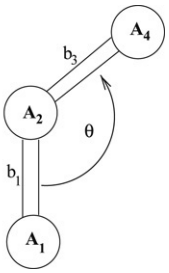
5. Application to protein angles

The backbone of a protein comprises a sequence of atoms

$$\mathrm{N}_1\text{–}\mathrm{C}_1^\alpha\text{–}\mathrm{C}_1\text{–}\mathrm{N}_2\text{–}\mathrm{C}_2^\alpha\text{–}\mathrm{C}_2\text{–}\cdots\text{–}\mathrm{N}_p\text{–}\mathrm{C}_p^\alpha\text{–}\mathrm{C}_p,$$

Table 2
For a sequence of atoms (A_1, A_2, A_3, A_4) as specified, with A_3 directly behind A_2, and A_1 directly below A_2, we label the angle shown in the sketch as one of ϕ, ψ, ω

θ	A_1	A_2	A_3	A_4
ϕ_i	C_{i-1}	N_i	C_i^α	C_i
ψ_i	N_i	C_i^α	C_i	N_{i+1}
ω_i	C_{i-1}^α	C_{i-1}	N_i	C_i^α



and by choosing 4 atoms with A_3 directly behind A_2, and A_1 directly below A_2 (see Table 2) we can specify 3 *dihedral angles*: ϕ, ψ, ω. The angle ω is restricted to be about zero, and is of little interest. The remaining angles (ϕ, ψ) are measured between $-\pi$ and π. A scatter plot of a collection of $(\phi_i, \psi_i), i = 1, \ldots, n$ for a protein is known as a *Ramachandran plot*, and has been used to characterize the secondary structure of the protein.

We can extend the result of Section 2 to multivariate data by using a multiplicative kernel, with equal bandwidths in each dimension. In two dimensions, if f is assumed to be a multivariate von Mises, with independent components, and common concentration κ, then we can approximate the asymptotic integrated variance of the kernel density estimate as $\nu/(4n\pi)$ with asymptotic integrated bias-squared as

$$\kappa^2 \left[3I_0(2\kappa)I_2(2\kappa) + I_1(2\kappa)^2\right]/(32\pi^2 I_0(\kappa)^4 \nu^2).$$

Hence in this case, the rule of thumb is

$$\nu = \left[n\hat{\kappa}^2 \left\{3I_0(2\hat{\kappa})I_2(2\hat{\kappa}) + I_1(2\hat{\kappa})^2\right\}/(4\pi I_0(\hat{\kappa})^4)\right]^{(1/3)}. \tag{10}$$

We illustrate a kernel density estimate for the protein *Malate dehydrogenase* which has $n = 343$ observations. The co-ordinates of the backbone can be obtained from the RCSB protein data bank (protein 5mdh); in turn the dihedral angles are easily calculated and the Ramachandran plot is shown in Fig. 2. For the purposes of this example, we assume that the sequence of angles is independent. To obtain $\hat{\kappa}$ we use the geometric mean of the estimated concentrations of the marginal data (using the wrapped estimate with $K = 3$). We obtain $\hat{\kappa} = 5.69$ and so, using (10), we use the smoothing parameter $\nu = 36.85$ in a multiplicative kernel. A contour plot of the square root – the transformation was used in order to reveal more details – of the estimated density is shown in Fig. 2. These density estimates can be used to suggest subpopulations within the data which can be compared with more standard representations of the secondary structure of the protein. Interesting further analysis is possible by comparing the density estimates of angles associated with various types of amino acids (which are defined by the side chains adjoining each C^α atom), but this is yet to be fully explored.

6. Concluding remarks

Extending some of the above results to a mixture of von Mises distributions would also be straightforward, and would proceed along the lines of Marron and Wand (1992). However, although we could obtain expressions for the approximate MISE, it would depend on the mixing proportions (as well as the means and concentrations of each component), and no plug-in rule would be readily available.

Agostinelli (2007) has considered alternative approaches to the robust estimation of κ which could also be used in (7) in place of those considered here. However, based on our simulation results, we would recommend trying both ν_1 and ν_3 as smoothing parameters. These are simple to compute, and in each distribution one of these performed well.

Finally, we note the survey paper of Jones et al. (1996) which addresses the issue of bandwidth selection for real-valued data. In addition to the ideas of the current paper, there are several alternatives which will have a counterpart for directional data. In particular, there are now well-known results in the Euclidean case which obtain more sophisticated

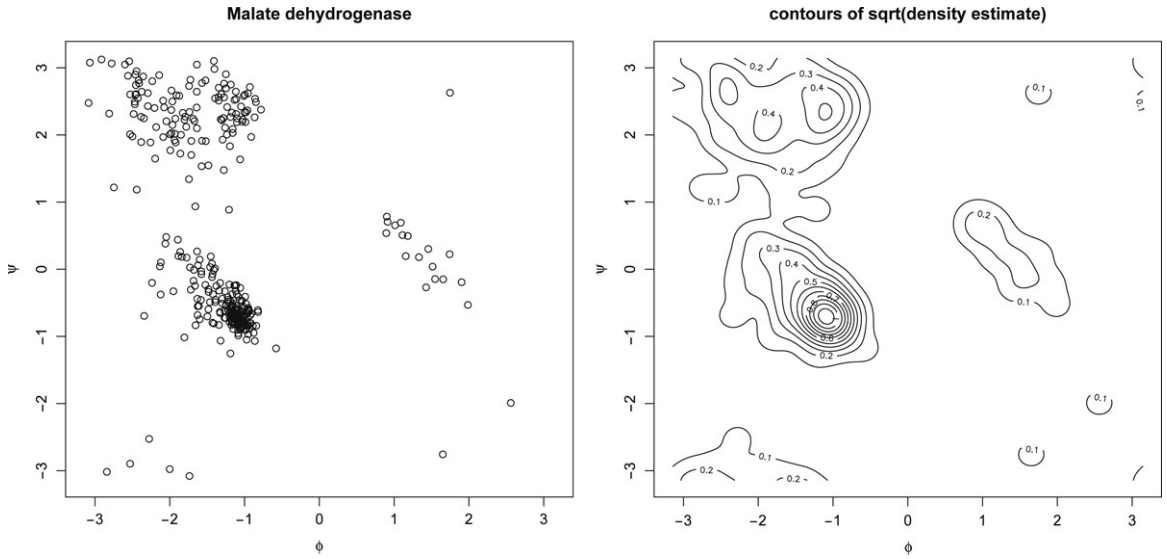


Fig. 2. Left: Ramachandran plot for Malate dehydrogenase, and Right: contour plot of the estimated (sqrt) density.

plug-in rules by estimating functionals of the derivatives. By using the results of Klemelä (2000) it should be possible to obtain circular data counterparts.

Acknowledgements

I would like to thank John Kent and Kanti Mardia for useful discussions.

References

Agostinelli, C., 2007. Robust estimation for circular data. Computational Statistics and Data Analysis 51, 5867–5875.
Fisher, N.I., 1989. Smoothing a sample of circular data. Journal of Structural Geology 11, 775–778.
Hall, P., Watson, G.S., Cabrera, J., 1987. Kernel density estimation with spherical data. Biometrika 74, 751–762.
Jammalamadaka, S. Rao, SenGupta, A., 2001. Topics in Circular Statistics. World Scientific, Singapore.
Jones, M.C., Marron, J.S., Sheather, S.J., 1996. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91, 401–407.
Klemelä, J., 2000. Estimation of densities and derivatives of densities with directional data. Journal of Multivariate Analysis 73, 18–40.
Ko, D., 1992. Robust estimation of the concentration parameter of the von-Mises distribution. The Annals of Statistics 20, 917–928.
Mardia, K.V., Jupp, P.E., 1999. Directional Statistics. John Wiley, Chichester.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. The Annals of Statistics 20, 712–736.
Ronchetti, E., 1992. Optimal robust estimators for the concentration parameter of a von Mises–Fisher distribution. In: Mardia, K.V. (Ed.), The Art of Statistical Science. A tribute to G.S. Watson. pp. 65–74.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.