No. 34

Theoretical Properties of Bandwidth Selectors for Kernel Density Estimation on the Circle

Yasuhito Tsuruta

Masahiko Sagae

9 June 2017

Theoretical Properties of Bandwidth Selectors for Kernel Density Estimation on the Circle

Yasuhito Tsuruta * Masahiko Sagae †

2017/6/09

Abstract

We derive the asymptotic properties of the least squares cross-validation (LSCV) selector and the direct plug-in rule (DPI) selector in the kernel density estimation for circular data. The DPI selector has the convergence rate $O(n^{-5/14})$, although the rate of the LSCV selector is $O(n^{-1/10})$. Our simulation shows that the DPI selector has a smaller variance and more stability than the LSCV selector, even when n is not large enough. In other words, the DPI selector outperforms the LSCV selector with

1 Introduction

Kernel density estimation is the standard nonparametric method for exploring the structure of circular data that distributes on the circle. The structure of the kernel density estimator is largely influenced by the value of the smoothing parameter. Therefore, the selection of the smoothing parameter is an important problem in the practical analysis of circular data.

Automatic selectors of the smoothing parameter for circular data are proposed, and the practical performances studied through the simulation. However, to our knowledge, no study has derived theoretical properties for the selectors in the field of circular data analysis.

We will explore the least squares cross-validation (LSCV) selector proposed by [Hall et al.(1987)] and the direct plug-in rule (DPI) selector proposed by [Di Marzio et al.(2011)]. The LSCV selector has been commonly used in the circular data analysis because of the simple definition. A few studies researched the properties of the DPI selector for circular data. However, in the studies on selectors for real line data, [Wand and Jones (1994)] pointed out that the DPI selector has a better performance than the LSCV selector with respect to the convergence rate to the optimal smoothing parameter.

This paper derives theoretical properties of both the LSCV selector and the DPI selector that include the asymptotic normality and the convergence rate: See [Hall and Marron(1987)], [Scott and Terrell (1987)], and [Sheather and Jones (1991)] for previous studies regarding real line data. The authors obtained the rates from the central limit theorem of a degenerate U-statistic given by [Hall(1984)]. We demonstrate that the converge rate of the DPI selector is $O(n^{-5/14})$ and that of the LSCV selector is $O(n^{-1/10})$. Numerical experiments show that DPI is much more stable than LSCV even when sample size n is not large enough.

*Graduate School of Human and Socio-Environmental Studies, Kanazawa University.
†School of Economics, Kanazawa University.

2 Properties of kernel density estimation

We give the definitions and the asymptotic properties of the kernel density estimators on the circle. A kernel density estimator $\hat{f}_\kappa(\theta)$ of unknown density f based on a random sample $\Theta_1, \ldots, \Theta_n$ is defined as

$$\hat{f}_\kappa(\theta) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(\theta - \Theta_i),$$

where $K_\kappa(\theta)$ is a symmetric kernel function, and κ is a concentration parameter that plays the role of a smoothing parameter, and corresponds to $\kappa = h^{-2}$ for a general bandwidth $h > 0$. Our loss function between $\hat{f}_\kappa$ and f is integrated squared error (ISE) given by $\mathrm{ISE}[\hat{f}_\kappa] := \int_{-\pi}^{\pi}\{\hat{f}_\kappa(\theta) - f(\theta)\}^2 d\theta$. The risk is mean integrated squared error (MISE) given by $\mathrm{MISE}[\hat{f}_\kappa] := \mathrm{E}_f[\mathrm{ISE}[\hat{f}_\kappa]]$.

We now employ a kernel function for circular data that is proposed by [Hall et al.(1987)].

Definition 1. (Kernel function)

A function $K_\kappa(\theta) : [-\pi, \pi) \to \mathbb{R}$ is said to be a kernel function. Let $K_\kappa(\theta)$ denote $K_\kappa(\theta) := C_\kappa^{-1}(L)L_\kappa(\theta)$, where

$$L_\kappa(\theta) := L(\kappa\{1 - \cos(\theta)\}) \tag{2.1}$$

and $C_\kappa(L) := \int_{-\pi}^{\pi} L_\kappa(\theta)d\theta$. We define the l th moment of L as

$$\mu_l(L) := \int_0^\infty L(r)r^{(l-1)/2}dr,$$

where $l \geq 0$ is even and $r = \kappa\{1 - \cos(\theta)\}$. The main term L satisfies the following seven conditions:

(a) The fourth derivative $L^{(4)}(r) := d^{(4)}L(r)/dr^4$ is continuous,

(b) If r is large, then, $L(r)r^{(p+1)/2} = O(r^{-(p+4)/2})$,

(c) The term $\delta_{2t}(L) := \int_{-\infty}^{\infty} L^2(z^2/2)z^{2t}dz$ has bounded for $t = 0, 1$.

(d) The moments $\mu_l(L)$ has bounded for $0 \leq l \leq p+4$, and $\mu_l(L) = \mu_{\kappa,l}(L) + O(\kappa^{-(p+6)/2})$, where $\mu_{\kappa,l}(L) := \int_0^\kappa L(r)r^{(l-1)/2}dr$.

(e) $\lim_{|z|\to\infty} \eta(z)|z|^{3/2} = o(1)$, where $\eta(z) := \int_{-\infty}^{\infty} L(t^2/2)L((t+z)^2/2)dt$.

(f) $\lim_{|z|\to\infty} \lambda(z)|z|^{3/2} = o(1)$, where $\lambda(L) := \int_{-\infty}^{\infty} L'(t^2/2)L((t+z)^2/2)t^2/2dt$ is bounded,

(g) The term $\delta_t(S_4^m) := \int_{-\infty}^{\infty} S_4^{2m}(z^2/2)z^{2t}dz$ is bounded for $t = 1, 2$ and $m = 1, 2$, where $S_4(z^2/2) := 3S^{(2)}(z^2/2) - 6z^2S^{(3)}(z^2/2) + z^4S^{(4)}(z^2/2)$.

The conditions (a), (c), and (d) are required to derive $\mathrm{MISE}[\hat{f}_\kappa]$; we can replace the condition (a) on the assumption that L' is continuous. We use the conditions (a)–(f) to prove the theoretical properties regarding the LSCV selector, and also need the conditions (a), (c), (d), and (g) to prove the theoretical properties regarding the DPI selector.

The kernel consisting of $L(r) = e^{-r}$ satisfies all the conditions of Definition 1, and it is equivalent to the von Mises (VM) kernel such as $L_\kappa(\theta) = \exp[-\kappa\{1 - \cos(\theta)\}]$. [Hall et al.(1987)] suggested that smooth and rapidly varying kernels of type (2.1) are asymptotically equivalent to the kernel of $L(r) = e^{-r}$.

We now define the pth-order kernel function.

Definition 2. (p th-order kernel function)
Let $p \geq 2$ be even. We say that $K_\kappa(\theta)$ is a pth-order kernel, if,

$$\mu_0(L) \neq 0, \qquad \mu_l(L) = 0, \quad l = 2, 4, \cdots, p-2, \qquad \text{and} \quad \mu_l(L) \neq 0 \quad l = p.$$

Let $R(g(\theta)\theta^t) := \int_{-\pi}^{\pi} g^2(\theta)\theta^{2t} d\theta$. Then, [Tsuruta and Sagae (2016)] derived the following asymptotic MISE by combining Lemma A.1–A.3 in Appendix A.

Theorem 1. Assume that the following conditions hold:

(i) $\kappa = \kappa(n)$ and $\lim_{n\to\infty} \kappa(n) = \infty$,

(ii) $\lim_{n\to\infty} n^{-1}\kappa^{1/2}(n) = 0$,

(iii) f is $p+2$th differentiable and $f^{(s)}$ is square-integrable for $s = 1, 2, \cdots, p$.

Then, employing a pth-order kernel, MISE is given by

$$\text{MISE}[\hat{f}_\kappa] = \frac{\mu_p^2(L)}{\mu_0^2(L)} R\left(\sum_{t=1}^{p/2} \frac{b_{p,2t} f^{(2t)}}{(2t)!}\right)\kappa^{-p} + n^{-1}\kappa^{1/2} d(L) + o(\kappa^{-p} + n^{-1}\kappa^{1/2}). \tag{2.2}$$

The details of the constants of $b_{p,2t}$ and $d(L)$ are presented in Appendix A. The first two terms on the left side of (2.2) are referred to AMISE$[\hat{f}_\kappa]$. When we employ a second-order kernel, we show that AMISE$[\hat{f}_\kappa]$ is equivalent to

$$\text{AMISE}[\hat{f}_\kappa] = \frac{\mu_2^2(L)}{\mu_0^2(L)} R(f'')\kappa^{-2} + n^{-1}\kappa^{1/2} d(L), \tag{2.3}$$

where $b_{2,2} = 2$, and the minimizer κ_* of (2.3) is given by

$$\kappa_* = \beta(L) R(f^{(2)})^{2/5} n^{2/5}, \tag{2.4}$$

where $\beta(L) := [4\mu_2^2(L)/\{\mu_0^2(L)d(L)\}]^{2/5}$.

The sketch of the proof is presented in Appendix-D in Tsuruta and Sagae (2016).

Higher-order kernels for $p \geq 4$ improve the MISE, but sacrifice the non-negative value so that the lower moments are 0. Therefore, we believe that most researchers prefer second-order kernels in practical analysis, because they are non-negative kernels. This paper focuses on the selectors of κ_*, which is the optimal smoothing parameter for second-order kernel density estimators.

3 Bandwidth selectors

3.1 Least squares cross-validation

The motivation of the LSCV selector comes from the minimization of ISE$[\hat{f}_\kappa] - R(f)$. The LSCV selector $\hat{\kappa}_{\text{CV}}$ is defined as the minimizer of the CV function given by

$$\text{CV}(\kappa) := R(\hat{f}) - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(\Theta_i), \tag{3.1}$$

where $\hat{f}_{-i}(\Theta_i) = (n-1)^{-1}\sum_{i\neq j}^{n} K_\kappa(\theta - \Theta_j)$. Hereafter, we use only the second-order kernels with respect to LSCV. If n is sufficiently large, then the equation (3.1) is replaced by

$$\mathrm{CV}(\kappa) := \frac{R(K_\kappa)}{n} + \frac{2}{n^2}\sum_{i<j}\gamma(y_{ij}), \tag{3.2}$$

where $y_{ij} := \Theta_i - \Theta_j$ and $\gamma(y) = \int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w+y)dy - 2K_\kappa(y)$. We apply the augmented cross-validation $\overline{\mathrm{CV}}(\kappa)$ given by

$$\overline{\mathrm{CV}}(\kappa) := \mathrm{CV}(\kappa) + \frac{2}{n}\sum_{i} f(\Theta_i) - R(f),$$

for theoretical analysis. Then, we obtain the variance of $\overline{\mathrm{CV}}(\kappa)$ that has a faster order than that of $CV(\kappa)$, and is similar to that derived by [Scott and Terrell (1987)]: they indicated that the augmented cross-validation for the real line data provides a smaller variance. We derive the expectation and variance of $\overline{\mathrm{CV}}(\kappa)$ as the following theorem.

Theorem 2. Assume that the three conditions of Theorem 1, $R(f^{(4)}f^{1/2}) < \infty$, and $R((f^{(4)})^{1/2}f) < \infty$.

Then, it follows that

$$\mathrm{E}_f[\overline{\mathrm{CV}}(\kappa)] = \mathrm{AMISE}[\hat{f}_\kappa] + o(\kappa^{-2} + n^{-1}\kappa^{1/2}), \tag{3.3}$$

and

$$\mathrm{Var}_f[\overline{\mathrm{CV}}(\kappa)] = \frac{2}{n^2}\kappa^{1/2}Q(L)R(f) + o(n^{-2}\kappa^{1/2} + n^{-1}\kappa^{-2}), \tag{3.4}$$

where $Q(L) := \int_{-\infty}^{\infty}\left\{2^{-1}\mu_0^{-2}(L)\eta(z) - 2^{1/2}\mu_0^{-1}(L)L(z^2/2)\right\}^2 dz$.

Proof. We set $\gamma(y_{ij}) = \gamma_{ij}$ to ease of notation. First, we calculate the expectation of $\overline{\mathrm{CV}}(\kappa)$, which is given by

$$\mathrm{E}_f[\overline{\mathrm{CV}}(\kappa)] = \frac{R(K_\kappa)}{n} + \frac{2}{n^2}\sum_{i<j}\mathrm{E}_f[\gamma_{ij}] + \frac{2}{n}\sum_{i}\mathrm{E}_f[f(\Theta_i)] - R(f). \tag{3.5}$$

We set $\gamma_i = \mathrm{E}_f[\gamma_{ij}|\Theta_i]$. Then, the conditional expectation γ_i is given by

$$\gamma_i = -f(\Theta_i) + f^{(4)}(\Theta_i)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}). \tag{3.6}$$

The details are presented in Appendix B. It follows from (3.6) that

$$\begin{aligned}\mathrm{E}_f[\gamma_{ij}] &= \mathrm{E}_f[\gamma_i] \\ &= -R(f) + R(f^{(2)})\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}).\end{aligned} \tag{3.7}$$

By considering Lemma A.3, (3.7), and $\mathrm{E}_f[f(\Theta_i)] = R(f)$, we obtain that $\mathrm{E}_f[\overline{\mathrm{CV}}(\kappa)]$ is equivalent to (3.3).

We calculate the variance of $\overline{\mathrm{CV}}(\kappa)$. That is,

$$\mathrm{Var}_f[\overline{\mathrm{CV}}(\kappa)] \simeq \frac{2}{n^2}\mathrm{Var}_f\left[\gamma_{ij}\right] + \frac{4}{n}\mathrm{Var}_f\left[f(\Theta_i)\right] + \frac{4}{n}\mathrm{Cov}_f\left[\gamma_{ij}, \gamma_{ik}\right] + \frac{8}{n}\mathrm{Cov}_f\left[\gamma_{ij}, f(\Theta_i)\right], \tag{3.8}$$

where $j \neq k$. Let $I_1 := R((f^{(4)})^{1/2}f)$, $I_2 := R(f^{(2)})R(f)$, and $I_{3:} = R(f^{3/2}) - R(f)^2$. Each term of the right side regarding (3.8) are given by

$$\mathrm{Var}_f[\gamma_{ij}] = \kappa^{1/2}[Q(L)R(f) + o(1)], \tag{3.9}$$

$$\mathrm{Var}_f[f(\Theta_i)] = I_3, \tag{3.10}$$

$$\mathrm{Cov}_f[\gamma_{ij}, \gamma_{ik}] = I_3 - 2\{I_1 - I_2\}\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}), \tag{3.11}$$

and

$$\mathrm{Cov}_f[\gamma_{ij}, f(\Theta_i)] = -I_3 + \{I_1 - I_2\}\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}). \tag{3.12}$$

The details of (3.9)–(3.10) are presented in Appendix C. By considering (3.8)– (3.10), we obtain that $\mathrm{Var}_f[\overline{\mathrm{CV}}(\kappa)]$ is equivalent to (3.4). □ □

With a strategy similar to [Scott and Terrell (1987)], Theorem 2 leads to that the LSCV selector $\hat{\kappa}_{\mathrm{CV}}$ is consistent with the minimizer κ_*.

Corollary 1. Let $\hat{\kappa}_{\mathrm{CV}} := \underset{\kappa \in (a\kappa_*, b\kappa_*)}{\arg\min}\ \mathrm{CV}(\kappa)$ for $a < b$. Then, it holds that

$$\hat{\kappa}_{\mathrm{CV}}/\kappa_* \xrightarrow{p} 1,$$

as $n \to \infty$.

Proof. We set $c := \hat{\kappa}_{\mathrm{CV}}/\kappa_*$. Then, it is derived from combining Theorem 1 and Theorem 2 that

$$\mathrm{AMISE}(c\kappa_*)/\mathrm{MISE}(c\kappa_*) \xrightarrow{p} 1, \tag{3.13}$$

$$\overline{\mathrm{CV}}(c\kappa_*)/\mathrm{MISE}(c\kappa_*) \xrightarrow{p} 1, \tag{3.14}$$

and

$$\mathrm{AMISE}(c\kappa_*)/\mathrm{AMISE}(\kappa_*) = \frac{1}{5c^2} + \frac{4c^{1/2}}{5}. \tag{3.15}$$

The equation (3.15) is the convex function such as the minimum at $c = 1$. Thus, if $c \neq 1$ and n is large, then it follows from combining (3.13) and (3.15) that

$$\mathrm{MISE}(c\kappa_*) > \mathrm{MISE}(\kappa_*). \tag{3.16}$$

Suppose that c does not converge to 1. Recall that it is necessary that $\overline{\mathrm{CV}}(c\kappa_*) \leq \overline{\mathrm{CV}}(\kappa)$ for any κ, because $\hat{\kappa}_{\mathrm{CV}}$ is the minimizer of $\overline{\mathrm{CV}}(\kappa)$. Also, if n is large, then it is shown that $\overline{\mathrm{CV}}(\kappa)$ is the convex function such as the minimum at $\kappa = c\kappa_*$, because we obtain that $\overline{\mathrm{CV}}(\kappa)$ approximates $\mathrm{AMISE}(\kappa)$ from Theorem 2. Therefore, it follows that

$$P(\overline{\mathrm{CV}}(c\kappa_*) < \overline{\mathrm{CV}}(\kappa_*)) \to 1, \tag{3.17}$$

as $n \to \infty$. From (3.14) and (3.17), then it holds that

$$\mathrm{MISE}(c\kappa_*) < \mathrm{MISE}(\kappa_*), \tag{3.18}$$

as $n \to \infty$. By contradiction between (3.16) and (3.18), this completes the proof. □ □

3.2 Direct plug-in rule

Note that $\psi_r := \int_{-\pi}^{\pi} f^{(r)}(\theta)f(\theta)d\theta$ and $R(f^{(r)}) = (-1)^r\psi_{2r}$. We now define the DPI estimator as

$$\hat{\kappa}_{\text{PI}} = \beta(L)\hat{\psi}_4(g)^{2/5}n^{2/5},$$

where

$$\hat{\psi}_4(g) := n^{-1}\sum_{i=1}^{n}\hat{f}_g^{(4)}(\Theta_i) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}T_g^{(4)}(\Theta_i - \Theta_j), \tag{3.19}$$

where $T_g^{(4)}(\theta) := C_\kappa^{-1}(L)S_g^{(4)}(\theta)$, and g and $T_g(\theta) := C_\kappa^{-1}(S)S_g(\theta)$ are a smoothing parameter and a kernel that is possibly different from κ and K_κ, respectively. The main term $S_g^{(4)}(\theta)$ is given by

$$\begin{aligned} S_g^{(4)}(\theta) := &-g\cos(\theta)S_g^{(1)}(\theta) + g^2\{-4\sin^2(\theta) + 3\cos^2(\theta)\}S_g^{(2)}(\theta) \\ &+ 6g^3\cos(\theta)\sin^2(\theta)S_g^{(3)}(\theta) + g^4\sin^4(\theta)S_g^{(4)}(\theta). \end{aligned} \tag{3.20}$$

The asymptotic properties for mean square error (MSE) of $\hat{\psi}_4$ play an important role in showing the theoretical properties of $\hat{\kappa}_{\text{PI}}$ in the next section. We provide the bias and the variance of $\hat{\psi}_4(g)$ in the following theorem.

Theorem 3. Assume that the following conditions hold:

(i) $g := g(n)$, $\lim_{n\to\infty} g(n) = \infty$, and $\lim_{n\to\infty} n^{-2}g^{9/2}(n) = 0$,

(ii) f is $(4+p)$th differentiable, ψ_{4+2t} is bounded for $t = 1, 2, \ldots, p/2$.

Then, when we employ a pth-order kernel, the bias is given by

$$\text{Bias}_f[\hat{\psi}_4(g)] = \text{Abias}_f[\hat{\psi}_4(g)] + O(n^{-1}g^{3/2} + g^{-(p+2)/2}), \tag{3.21}$$

where,

$$\text{Abias}_f[\hat{\psi}_4(g)] = \frac{3g^{5/2}S_g^{(2)}(0)}{2^{1/2}\mu_0(S)n} + \frac{\mu_p(S)}{\mu_0(S)}\sum_{t=1}^{p/2}\frac{b_{p,2t}\psi_{4+2t}}{(2t)!}g^{-p/2},$$

and the variance is given by

$$\text{Var}_f[\hat{\psi}_4(g)] = \frac{4}{n}\text{Var}[f^{(4)}(\Theta_i)] + \frac{2G_{1,0}(S_4)\psi_0 g^{9/2}}{n^2} + o(n^{-1} + n^{-2}g^{9/2}). \tag{3.22}$$

where $G_{m,t}(S_4) := 2^{-m}\mu_0^{-2m}(S)\delta_t(S_4^m)$.

Proof. Let $U_{ij} = T_g^{(4)}(\Theta_i - \Theta_j)$, and$U_i = \text{E}_f[U_{ij}|\Theta_i]$. The expectation of $\hat{\psi}_4(g)$ is given by

$$\text{E}_f[\hat{\psi}_4(g)] = n^{-1}T_g^{(4)}(0) + 2n^{-2}\sum_{i<j}\text{E}_f[U_{ij}]. \tag{3.23}$$

It follows from (3.20) that

$$S_g^{(4)}(0) = 3g^2[S_g^{(2)}(0) + O(g^{-1})]. \tag{3.24}$$

By combining (3.24) and Lemma A.1, we obtain that the first term of the right side of (3.23) is equal to

$$n^{-1}T_g^{(4)}(0) = \frac{3g^{5/2}[S_g^{(2)}(0) + O(g^{-1})]}{2^{1/2}\mu_0(S)n}. \tag{3.25}$$

It follows from Lemma A. 2 that

$$\begin{aligned}
U_i &= \int_{-\pi}^{\pi} T_g^{(4)}(\theta_j - \Theta_i) f(\theta_j) d\theta_j \\
&= \int_{-\pi}^{\pi} T_g(\theta_j - \Theta_i) f^{(4)}(\theta_j) d\theta_j \\
&= \sum_{t=0}^{p/2} \frac{f^{(4+2t)}(\Theta_i)}{(2t)!} \alpha_{2t}(T_g) + O(\alpha_{p+2}(T_g)) \\
&= f^{(4)}(\Theta_i) + \mu_0^{-1}(S)\mu_p(S) g^{-p/2} \sum_{t=1}^{p/2} \frac{b_{p,2t} f^{(4+2t)}(\Theta_i)}{(2t)!} + O(g^{-(p+2)/2}),
\end{aligned} \tag{3.26}$$

The expectation $\mathrm{E}_f[U_{ij}]$ of (3.23) is given by the expectation of (3.26) over Θ_i

$$\begin{aligned}
\mathrm{E}_f[U_{ij}] &= \mathrm{E}_f[U_i] \\
&= \psi_4 + \mu_0^{-1}(S)\mu_p(S) g^{-p/2} \sum_{t=1}^{p/2} \frac{b_{p,2t}\psi_{4+2t}}{(2t)!} + O(g^{-(p+2)/2}).
\end{aligned} \tag{3.27}$$

We obtain the bias (3.21) from combining (3.23), (3.25) and (3.27).

We derive the variance of $\hat{\psi}_4(g)$. We set $W_{ij} := U_{ij} - U_i - U_j + \mathrm{E}_f[U_i]$ and $Z_i := U_i - \mathrm{E}_f[U_i]$. Then, we obtain that $\mathrm{E}_f[W_{ij}] = 0$, $\mathrm{E}_f[Z_i] = 0$ and $\mathrm{Cov}_f[Z_i W_{ij}] = 0$. By using W_{ij} and Z_i. we present $\hat{\psi}_4(g) - \mathrm{E}_f[\hat{\psi}_4(g)]$ as

$$\hat{\psi}_4(g) - \mathrm{E}_f[\hat{\psi}_4(g)] = \frac{2(n-1)}{n^2} \sum_i Z_i + \frac{2}{n^2} \sum_{i<j} W_{ij}. \tag{3.28}$$

Thus, the variance of $\hat{\psi}_4$ is equal to

$$\begin{aligned}
\mathrm{Var}_f[\hat{\psi}_4(g)] &= \mathrm{E}_f \left[\left\{ \frac{2(n-1)}{n^2} \sum_i Z_i + \frac{2}{n^2} \sum_{i<j} W_{ij} \right\}^2 \right] \\
&= \frac{4(n-1)^2}{n^4} \sum_i \mathrm{Var}_f[Z_i] + \frac{4}{n^4} \sum_{i<j} \mathrm{Var}_f[W_{ij}].
\end{aligned} \tag{3.29}$$

By combining (3.26) and (3.27), $\mathrm{Var}_f[Z_i]$ is reduced to

$$\begin{aligned}\mathrm{Var}_f[Z_i] &= \mathrm{E}[U_i^2] - \mathrm{E}[U_i]^2 \\ &= \int_{-\pi}^{\pi} f^{(4)}(\theta_i)^2 f(\theta_i) d\theta_i - \left[\int_{-\pi}^{\pi} f^{(4)}(\theta_i) f(\theta_i) d\theta_i\right]^2 + o(1) \\ &= \mathrm{Var}[f^{(4)}(\Theta_i)] + o(1). \end{aligned} \tag{3.30}$$

By considering (3.27), $\mathrm{E}_f[U_{ij}^2] = g^{9/2}[G_{1,0}(S_4)\psi_0 + o(1)]$, and $\mathrm{E}[U_i^2] = \mathrm{E}[U_i]^2 = O(1)$ (The details of $\mathrm{E}_f[U_{ij}^2]$ and $\mathrm{E}[U_i^2]$ are presented in Appendix D.), we obtain $\mathrm{Var}_f[W_{ij}]$. That is,

$$\begin{aligned}\mathrm{Var}_f[W_{ij}] &= \mathrm{E}_f[U_{ij}^2] - 2\mathrm{E}_f[U_i^2] + \mathrm{E}_f[U_i]^2 \\ &= g^{9/2}[G_{1,0}(S_4)\psi_0 + o(1)]. \end{aligned} \tag{3.31}$$

We obtain (3.22) from combining (3.29) (3.30), and (3.31). □ □

Theorem 3 easily shows the optimal MSE.

Corollary 2. Select the optimal smoothing parameter $g_* > 0$ such as that $\mathrm{Abias}_f[\hat{\psi}(g)] = 0$. Then, g_* is given by

$$g_* = W(S) n^{2/(p+5)}, \tag{3.32}$$

where $W(S) = \left[-\{2^{1/2}\mu_p(S)\sum_{t=1}^{p/2}[\psi_{4+2t} b_{p,2t}/(2t)!]\}/\{3S_g^{(2)}(0)\}\right]^{2/(p+5)}$. The remaining squared bias can be ignored by $\mathrm{Bias}_f^2[\hat{\psi}_4(g)] = O(n^{-(2p+4)/(p+5)})$. In other words, The convergence rate of the minimum MSE depends on only the variance $\mathrm{Var}_f[\hat{\psi}_4(g)]$. If $p < 4$, then, $\inf_{g>0} \mathrm{MSE}[\hat{\psi}_r(g)]$ is equivalent to the second term of the right side of (3.22), if $p = 4$, then, it is equivalent to the first two terms of that, otherwise, it is equivalent to only the first term of that. Thus, $\inf_{g>0} \mathrm{MSE}[\hat{\psi}_4(g)]$ is presented as

$$\inf_{g>0} \mathrm{MSE}[\hat{\psi}_4(g)] = \begin{cases} O(n^{-(2p+1)/(p+5)}) & p < 4, \\ O(n^{-1}) & p \geq 4. \end{cases}$$

Corollary 2 indicates that employing a higher-order kernel for $p \geq 4$ achieves the parametric rate of $O(n^{-1})$. However, it is greatly difficult to inspect whichever higher-order kernels satisfy the positive condition of g_*, because the sign of g_* depends on T_g and the sum of some unknown functionals ψ_r; we may need to know the unknown density f to know the sign of g_*.

If T_g is a second-order kernel, then the sign of g_* depends on only T_g. Hence, we always obtain a positive g_* by applying the suitable kernels such that $\mu_2/S(0)$ is positive, because $\psi_6 = -R(f^{(3)})$. We recommend employing suitable second-order kernels such as a VM kernel to estimate ψ_4.

We obtain $g_* = [c\psi_6 n]^{2/7}$ for the suitable second-order kernels, where $c = -2^{-1/2}\mu_2(S)/(6S_g^{(2)}(0))$. Estimating g_* also requires estimating an unknown functional ψ_6. We provide the simplest estimator of ψ_6 by assuming that a true density is a VM density $f_{\mathrm{VM}}(\theta;\tau) := (2\pi I_0(\tau)^{-1}\exp\{\tau\cos(\theta)\}$, where $I_p(\tau)$ denotes the modified Bessel function of the first kind and the order p, and τ is the concentration parameter. The estimator of ψ_6 is given by

$$\hat{\psi}_6^{\mathrm{VM}} := -[4\hat{\tau}I_1(2\hat{\tau}) + 30\hat{\tau}^2 I_2(2\hat{\tau}) + 15\hat{\tau}^3 I_3(2\hat{\tau})]/\{16\pi I_0^2(\hat{\tau})\}.$$

We propose the easy and practical algorithm for direct plug-in rule, called for "One-step direct plug-in rule".

Algorithm 1. One-step direct plug-in rule conducts the following procedure:

Step.1 Calculate ML estimator $\hat{\tau}$ and $\hat{\psi}_6^{\mathrm{VM}}$.

Step.2 Give $\hat{g} := [c\hat{\psi}_6^{\mathrm{VM}} n]^{2/7}$ as the estimator of g_*.

Step.3 Give $\hat{\kappa}_{\mathrm{PI}} = \beta(L)\hat{\psi}_4(\hat{g})^{2/5} n^{2/5}$.

4 Theoretical properties for the selectors

From theoretical perspective, we must inspect whether the DPI selector outperforms the LSCV selector. The theoretical performance for the selector $\hat{\kappa}$ is measured by the convergence rate of the relative error: $\hat{\kappa}/\kappa_* - 1$. The rate is derived through the asymptotically normal distribution:

$$n^{\alpha}(\hat{\kappa}/\kappa_* - 1) \xrightarrow{d} N\left(0, \sigma^2\right),$$

where $\sigma^2 < \infty$ depends only on f and L, but not on n. The asymptotic distribution of LSCV and DPI are shown in Theorem 4 and Theorem 5, respectively.

Theorem 4. Assume that all the conditions of Theorem 1 and Theorem 2 hold: then, it holds that

$$n^{1/10}(\hat{\kappa}_{\mathrm{CV}}/\kappa_* - 1) \xrightarrow{d} N\left(0, \sigma_{\mathrm{CV}}^2\right), \tag{4.1}$$

as $n \to \infty$, where $\sigma_{\mathrm{CV}}^2 := 50 d^{-2}(L) M_{1,0}(L) R(f) \beta^{-1/2}(L) R(f'')^{-1/5}$, and $M_{m,t}(L) := \int_{-\infty}^{\infty} m(L)^{2m} z^{2t} dm$, where

$$m(L) := 2^{-1}\mu_0^{-2}(L)\{\eta(z) + \lambda(z) + \lambda(-z)\} - 2^{-1/2}\mu_0^{-1}(L)\{L(z^2/2) + L(z^2/2)z^2\}.$$

Theorem 5. Assume that the conditions of Theorem 3 hold. Then, when we employ the suitable second-order kernel, and it holds that

$$n^{5/14}(\hat{\kappa}_{\mathrm{PI}}/\kappa_* - 1) \xrightarrow{d} N\left(0, \sigma_{\mathrm{CV}}^2\right), \tag{4.2}$$

as $n \to \infty$, where $\sigma_{\mathrm{PI}}^2 = 8W^{9/2}(S) G_{1,0}(S_4) \psi_0 \psi_4^{-2}/25$.

Theorems 4 and 5 are proved by the asymptotic normality of a degenerate U-statistic given by [Hall(1984)]. We give the definition of a degenerate U-statistic. A U-statistic is defined as $U_n := \sum_{i<j} H_{ij}$, where $H_{ij} := H(\Theta_i, \Theta_j)$ and H_{ij} is symmetric and $\mathrm{E}_f[H_{ij}] = 0$. Let the degenerate U-statistic be the U-statistic satisfying $\mathrm{E}_f[H_{ij}|\Theta_i] = 0$. The following lemma describes the asymptotic normality of a degenerate U-statistic.

Lemma 1 ([Hall(1984)])**.** Assume that H_{ij} is symmetric, and $\mathrm{E}_f[H_{ij}|\Theta_i] = 0$, almost surely and $\mathrm{E}_f[H_{ij}^2 \Theta_i] < \infty$ for each n. We set $G_{ij} := \mathrm{E}[H_{ii} H_{ij}]$. if

$$\frac{\mathrm{E}[G_{ij}^2] + n^{-1}\mathrm{E}_f[H_{ij}^4]}{\mathrm{E}_f[H_{ij}^2]^2} \to 0, \tag{4.3}$$

as $n \to \infty$, then, it holds that

$$\sum_{1 \le i < j \le n} H_{ij} \xrightarrow{d} N(0, n^2 \mathrm{E}[H_{ij}^2]/2).$$

We now prove Theorem 4.

Proof of Theorem 4. If n is large, it follows from Lemma A.3 that

$$\mathrm{CV}(\kappa) \simeq \frac{d(L)\kappa^{1/2}}{n} + \frac{2}{n^2}\sum_{i<j}\gamma(y_{ij}). \tag{4.4}$$

The derivative of (4.4) is given by

$$\frac{d\mathrm{CV}(\kappa)}{d\kappa} \simeq \frac{d(L)}{2n\kappa^{1/2}} + \frac{2}{n^2\kappa^{1/2}}\sum_{i<j}V_{ij}, \tag{4.5}$$

where

$$\begin{aligned}
V_{ij} &:= \kappa^{-1/2}[\gamma(y_{ij}) + \rho(y_{ij}) + 3/4\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\tau(y_{ij})],\\
\phi_\kappa(y_{ij}) &:= \kappa C_\kappa^{-1}(L)\frac{d}{d\kappa}L_\kappa(y_{ij}),\\
\rho(y_{ij}) &:= K_\kappa(y_{ij}) + \int_{-\pi}^{\pi}\{\phi_\kappa(w)K_\kappa(w+y_{ij}) + K_\kappa(w)\phi_\kappa(w+y_{ij})\}dw - 2\phi_\kappa(y_{ij}),
\end{aligned}$$

and

$$\tau(y_{ij}) := \int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w+y_{ij})dw - K_\kappa(y_{ij}).$$

The details are presented in Appendix E. The selector $\hat{\kappa}_{\mathrm{CV}}$ satisfies that $d\mathrm{CV}(\kappa)/d\kappa\mid_{\kappa=\hat{\kappa}_{\mathrm{CV}}}= 0$. This is equivalent to

$$2n^{-2}\sum_{i<j}V_{ij}\Bigg|_{\kappa=\hat{\kappa}_{\mathrm{CV}}} = -d(L)/(2n). \tag{4.6}$$

Note that $V_i := \mathrm{E}_f[V_{ij}|\Theta_i]$. Then, we set $H_{ij} := V_{ij} - V_i - V_j + \mathrm{E}_f[V_i]$ and $X_i := V_i - \mathrm{E}_f[V_i]$. Then, we rewrite $2n^{-2}\sum_{i<j}\{V_{ij} - \mathrm{E}_f[V_{ij}]\}$ as

$$2n^{-2}\sum_{i<j}V_{ij} - 2n^{-2}\sum_{i<j}\mathrm{E}_f[V_{ij}] \simeq 2n^{-1}\sum_i X_i + 2n^{-2}\sum_{i<j}H_{ij},$$

where $2n^{-2}\sum_{i<j}H_{ij}$ is the degenerate U-statistic. We obtain the asymptotic normality for $2n^{-1}\sum_i X_i$ from the standard central limit theorem (CLT). That is,

$$\frac{2}{n}\sum_i X_i \xrightarrow{d} N\left(0, Bn^{-1}\kappa^{-5}\right), \tag{4.7}$$

where, $B := 16\mu_2^4(L)\{R(f^{(4)}f^{1/2}) - R(f'')^2\}/\{\mu_0^4(L)\}$. The details are presented in Appendix F. We obtain the asymptotic normality for $2n^{-2}\sum_{i<j}H_{ij}$ from Lemma 1. that is,

$$\frac{2}{n^2}\sum_{i<j}H_{ij} \xrightarrow{d} N(0, 2n^{-2}\kappa^{-1/2}M_{1,0}(L)R(f)). \tag{4.8}$$

See Appendix-G for details. It is derived from combining (4.7) and (4.8) that the asymptotically normal for $2n^{-2}\sum_{i<j} V_{ij}$ is

$$\frac{2}{n^2}\sum_{i<j} V_{ij} \xrightarrow{d} N\left(-2R(f'')\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-5/2}, \quad Bn^{-1}\kappa^{-5} + 2n^{-2}\kappa^{-1/2}M_{1,0}(L)R(f)\right). \tag{4.9}$$

We take $\kappa = \hat{\kappa}_{\mathrm{CV}}$ in (4.9). Then, we replace $\hat{\kappa}_{\mathrm{CV}}$ in the variance to κ_* by Corollary 1. Thus, it follows from combining (4.6) and (4.9) that

$$-2R(f'')\mu_0^{-2}(L)\mu_2^2(L)\hat{\kappa}_{\mathrm{CV}}^{-5/2} \xrightarrow{d} N\left(-\frac{d(L)}{2n}, \quad Bn^{-1}\kappa_*^{-5} + 2n^{-2}\kappa_*^{-1/2}M_{1,0}(L)R(f)\right). \tag{4.10}$$

the first term for the variance of (4.10) is ignored, because the convergence rate of the first term is $O(n^{-3})$, and that of the second term is $O(n^{-11/5})$ by using $\kappa_* = O(n^{2/5})$. From (2.4), we obtain that $R(f'')\mu_2^2(L)n/(d(L)\mu_0(L)) = \kappa_*^{5/2}$. Thus, (4.10) is reduced to

$$(\hat{\kappa}_{\mathrm{CV}}/\kappa_*)^{-5/2} \xrightarrow{d} N\left(1, \quad 8d(L)^{-2}M_{1,0}(L)R(f)\kappa_*^{1/2}\right). \tag{4.11}$$

Let $g(x) = x^{-5/2}$. Then, it follows that g(1)=1 and $\{g'(1)\}^2 = 25/4$. We obtain the asymptotic normality for $\hat{\kappa}_{\mathrm{CV}}/\kappa_*$ by applying the delta method to (4.11). That is,

$$\hat{\kappa}_{\mathrm{CV}}/\kappa_* \xrightarrow{d} N\left(1, 50d(L)^{-2}M_{1,0}(L)R(f)\beta(L)^{-1/2}R(f'')^{-1/5}n^{-1/5}.\right). \tag{4.12}$$

Theorem 4 completes the proof from (4.12). □ □

Next, we prove Theorem 5.

Proof of Theorem 5. The Taylor expansion $\hat{\kappa}_{\mathrm{PI}} = \hat{\kappa}_{\mathrm{PI}}(\hat{\psi}_4(g_*))$ is given by

$$\begin{aligned}\hat{\kappa}_{\mathrm{PI}}(\hat{\psi}_4(g_*)) &\simeq \beta(L)n^{2/5}\psi_4^{2/5} + \frac{2}{5}\beta(L)n^{2/5}\psi_4^{-3/5}(\hat{\psi}_4(g_*) - \psi_4)\\ &= \kappa_*[1 + 2(\hat{\psi}_4(g_*) - \psi_4)/(5\psi_4)].\end{aligned} \tag{4.13}$$

Equation (4.13) is reduced to

$$\hat{\kappa}_{\mathrm{PI}}/\kappa_* - 1 = \frac{2}{5\psi_4}(\hat{\psi}_4(g_*) - \psi_4). \tag{4.14}$$

Noting $W_{ij} := U_{ij} - U_i - U_j + \mathrm{E}_f[U_i]$, and $Z_i := U_i - \mathrm{E}_f[U_i]$, it follows that (3.28) becomes

$$\hat{\psi}_4(g) - \mathrm{E}_f[\hat{\psi}_4(g)] \simeq 2n^{-1}\sum_i Z_i + 2n^{-2}\sum_{i<j} W_{ij}, \tag{4.15}$$

where $2n^{-2}\sum_{i<j} W_{ij}$ is the degenerate U-statistic. From (3.30), we obtain the asymptotic normality distribution from the standard CLT. That is,

$$n^{-1/2}\sum_i Z_i \xrightarrow{d} N(0, \mathrm{Var}_f[f(\Theta_i)]). \tag{4.16}$$

If we choose $g_* = W(S)n^{2/7}$, then applying Lemma 1 to $2n^{-2}\sum_{i<j} W_{ij}$, it is given by

$$\frac{2}{n^2}\sum_{i<j} W_{ij} \xrightarrow{d} N(0, 2n^{-2}g_*^{9/2}G_{1,0}(S_4)\psi_0), \tag{4.17}$$

as $n \to \infty$. The details are presented in Appendix H. By combining (4.16) and (4.17), we obtain the asymptotic distribution of (4.15). That is,

$$\hat{\psi}_4(g_*) - \mathrm{E}_f[\hat{\psi}_4(g_*)] \xrightarrow{d} N(0, 4n^{-1}\mathrm{Var}_f[f(\Theta_i)] + 2n^{-2}g_*^{9/2}G_{1,0}(S_4)\psi_0). \tag{4.18}$$

Corollary 2 shows that the rate of $\mathrm{Var}_f[\hat{\psi}_4(g^*)]$ is the order $n^{-5/7}$. Thus, the equation (4.18) is reduced to

$$n^{5/14}\{\hat{\psi}_4(g_*) - \mathrm{E}_f[\hat{\psi}_4(g_*)]\} \xrightarrow{d} N(0, 2W^{9/2}(S)G_{1,0}(S_4)\psi_0). \tag{4.19}$$

The main team $\hat{\psi}_4(g_*) - \psi_4$ of the right side for (4.14) is equivalent to

$$n^{5/14}\{\hat{\psi}_4(g_*) - \psi_4\} = n^{5/14}\{\hat{\psi}_4(g_*) - \mathrm{E}_f[\hat{\psi}_4(g_*)]\} - n^{5/14}\mathrm{Bias}_f[\hat{\psi}_4(g_*)]. \tag{4.20}$$

We show that $\mathrm{Bias}_f[\hat{\psi}_4(g^*)] = O(n^{-4/7})$ from Corollary 2. Then, we obtain that $n^{5/14}\mathrm{Bias}_f[\hat{\psi}_4(g_*)]$ is $O(n^{-3/14})$. Thus, if n is large, then this term is ignored. Therefore, the asymptotic normal distribution for $n^{5/14}\{\hat{\psi}_4(g_*) - \psi_4\}$ is given by

$$n^{5/14}\{\hat{\psi}_4(g) - \psi_4\} \xrightarrow{d} N(0, 2W^{9/2}(S)G_{1,0}(S_4)\psi_0). \tag{4.21}$$

Therefore, as $n \to \infty$, Theorem 5 completes the proof from (4.21) and (4.14). □ □

The convergence rate of $\hat{\kappa}_{\mathrm{CV}}$ and $\hat{\kappa}_{\mathrm{PI}}$ are equivalent to that of LSCV selector and DPI selector on the real line, respectively ([Hall and Marron(1987)], [Scott and Terrell (1987)], and [Sheather and Jones (1991)]). The rate of $\hat{\kappa}_{\mathrm{PI}}$ is greatly faster than that of $\hat{\kappa}_{\mathrm{CV}}$. Moreover, $\hat{\kappa}_{\mathrm{PI}}$ is more stable with the smaller order of variance. Therefore, the DPI selector is more appealing with respect to theoretical performances than the LSCV selector.

5 Numerical experiment

Analyzing practical data of a small sample size often does not have the same effect as the theoretical results. Therefore, we need to perform a simulation for comparing the LSCV selector and the DPI selector. Our simulation in statistical software R is conducted by the following procedure:

1. Generate the random sample of size n distributed as the VM distribution $f_{\mathrm{VM}}(\theta; \tau = 1)$.

2. Calculate the optimal parameter κ_* applying $f_{\mathrm{VM}}(\theta; \tau = 1)$ to (2.4).

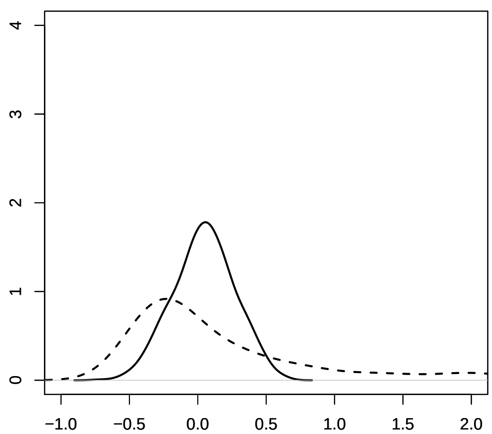
3. Estimate $\hat{\kappa}_{\mathrm{CV}}$ by bw.cv.mse.circular, which is the function in circular library of R.

4. Estimate $\hat{\kappa}_{\mathrm{PI}}$ by the One-step plug-in rule.

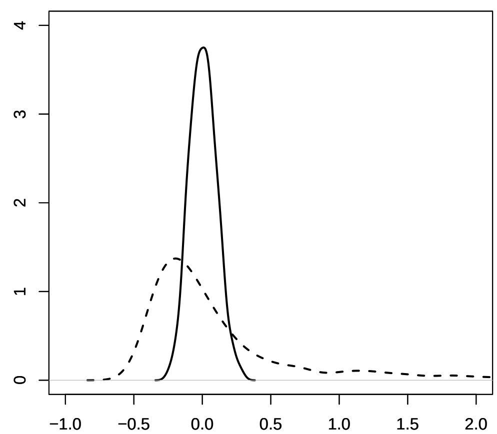
5. Calculate $Y_{\mathrm{CV}} = \log(\hat{\kappa}_{\mathrm{CV}}/k_*)$ and $Y_{\mathrm{PI}} = \log(\hat{\kappa}_{\mathrm{PI}}/k_*)$.

6. Repeat Step 1–5 1000 times, and give the kernel density estimators of Y_{CV} and Y_{PI} to estimate the bandwidth h with normal reference rule, respectively.

Fig. 1 shows that the DPI selector $\hat{\kappa}_{\text{PI}}$ has a smaller variance and bias than the LSCV selector $\hat{\kappa}_{\text{CV}}$. In other words, $\hat{\kappa}_{\text{PI}}$ is more stable. The LSCV selector $\hat{\kappa}_{\text{CV}}$ trends to be undersmoothing, although it is sometimes greatly oversmoothing even if $n = 1000$. The simulation indicates that $\hat{\kappa}_{\text{PI}}$ performs better than $\hat{\kappa}_{\text{CV}}$ in practical analysis.



(a) Sample size n is 100.



(b) Sample size n is 1000.

Figure 1: The solid line is the kernel density estimator of $\log(\hat{\kappa}_{\text{PI}}/k_*)$ and the broken line is the kernel density estimator of $\log(\hat{\kappa}_{\text{CV}}/k_*)$, where $\hat{\kappa}_{\text{PI}}$ is the DPI selector, $\hat{\kappa}_{\text{CV}}$ is the LSCV selector, and κ_* is the optimal smoothing parameter. The selectors are based on 1000 simulated samples of size $n = 100$ and 1000 from the VM density.

6 Conclusion

We derived the asymptotic properties for the least squares cross-validation selector and the direct plug-in selector for circular data. The convergence rate of the DPI selector is $O(n^{-5/14})$ and that of the LSCV selector is $O(n^{-1/10})$. The rates are equivalent to the two selectors on the real line, respectively. Thus, the theoretical performance of the DPI selector is better than that of the LSCV selector. Our simulation shows that the DPI selector is more stable than the LSCV selector.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers JP16K00043, JP24500339, and JP16H02790.

References

[Di Marzio et al.(2009)] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters* 79, 2066–2075.

[Di Marzio et al.(2011)] Di Marzio, M., Panzera, A. and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference* 141, 2156–2173.

[Feller(1968)] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. I (Third ed.)* (pp.244). USA: John Wiley & Sons, Inc.

[Hall(1984)] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, 14(1), 1–16.

[Hall and Marron(1987)] Hall, P., and Marron, J. S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields*, 74, 567–581.

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

[Scott and Terrell (1987)] Scott, D. W., and Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82, 1131–1146.

[Scott (1992)] Scott, David W. (1992) *Multivariate density estimation: theory, practice, and visualization (First ed.)*. USA: John Wiley & Sons.

[Sheather and Jones (1991)] Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683–690.

[Tsuruta and Sagae (2016)] Tsuruta, Y., and Sagae, M. , 2016. Higher order kernel density estimation on the circle. *Discussion Paper, No.30 in school of Economics, Kanazawa University.*

[Wand and Jones (1994)] Wand, M. P., and Jones, M. C. (1994). Bandwidth selection. *kernel smoothing* (pp.58–88). USA, Chapman&Hall/CRC.

Appendix A

Lemma A. 1. If K_κ is a pth-order kernel, then, $C_\kappa(L)$ is reduced to

$$C_\kappa(L) = \kappa^{-1/2}2^{1/2}\mu_0(L) + O(\kappa^{-(p+1)/2}).$$

Lemma A. 2. We set $\alpha_j(K_\kappa) := \int_{-\pi}^{\pi} K_\kappa(\theta)\theta^j d\theta$, $g_j(r/\kappa) := \{2 - r/\kappa\}^{(j-1)/2}$ for $j \geq 0$, and $a_s := (2s-2)!!/\{(2s-1)!!s\}$. the tth power of θ is given by

$$\theta^{2t} = \sum_{q=t}^{z/2} A_q(z,t)\{r/\kappa(2 - r/\kappa)\}^q + O(\kappa^{-(z+2)/2}), \quad 0 \leq \theta < \pi/2, \tag{A.1}$$

where

$$A_q(z,t) := \sum_{\sum_{s=1}^{z/2} t_s = t, \quad \sum_{s}^{z/2} s t_s = q} \frac{t!}{t_1! t_2! \cdots t_{z/2}!} \prod_{l=1}^{z/2} a_l^{t_l}. \tag{A.2}$$

Therefore, $\alpha_{2t}(K_\kappa)$ for even $2t \le z \le p+4$ is given by

$$\alpha_{2t}(K_\kappa) = 2C_\kappa^{-1}(L)\kappa^{-1/2} \sum_{q=t}^{z/2} \sum_{m=0}^{z/2-q} \kappa^{-(q+m)} A_q(z,t)(m!)^{-1} g_{2q}^{(m)}(0)\mu_{2(q+m)}(L) + O(\kappa^{-(z+2)/2}) \tag{A.3}$$

If K_κ is a pth-order kernel, then (A.3) is reduced to

$$\alpha_{2t}(K_\kappa) = b_{p,2t}\mu_0^{-1}(L)\mu_p(L)\kappa^{-p/2} + O(\kappa^{-(p+2)/2}) \quad 0 < j \le p, \tag{A.4}$$

where,

$$b_{p,2t} = 2^{1/2} \sum_{q=t}^{p/2} A_q(p,t)(\{p/2-q\}!)^{-1} g_{2q}^{(p/2-q)}(0).$$

Especially, the term $b_{2,2}$ is 2. It follows from (A.3) that

$$\alpha_{p+2}(K_\kappa) = O(\kappa^{-(p+2)/2}). \tag{A.5}$$

Lemma A. 3. The term $R(K(\theta)\theta^t)$ is equivalent to

$$R(K(\theta)\theta^t) := \kappa^{-(2t-1)/2}[d_{2t}(L) + o(1)],$$

where $d_{2t}(L) := 2^{-1}\mu_0^{-2}(L)\delta_{2t}(L)$ and $d(L) := d_0(L)$.

Appendix B

We will derive the conditional expectation γ_i. If t is odd, then, we obtain that the term $\int_{-\pi}^{\pi} \gamma(y)y^t dy = 0$, because the function $\gamma(y)$ is symmetry. By the binormal theorem, $\int_{-\pi}^{\pi} \gamma(y)y^{2t} dy$ is reduced to

$$\begin{aligned} \int_{-\pi}^{\pi} \gamma(y)y^{2t} dy &= \int K_\kappa(w) \int K_\kappa(s)(s-w)^{2t} dw ds - 2\alpha_{2t}(K_\kappa) \\ &= \sum_{m=0}^{2t} (-1)^m {}_{2t}C_m \alpha_m(K_\kappa)\alpha_{2t-m}(K_\kappa) - 2\alpha_{2t}(K_\kappa). \end{aligned} \tag{B.1}$$

Recalling that the kernel K_κ is second-order, by combining (B.1), (A.4), and (A.5), it is derived that

$$\int_{-\pi}^{\pi} \gamma(y)y^{2t} dy = \begin{cases} -1 & t = 0, \\ 0 & t = 1, \\ 24\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}) & t = 2, \\ O(\kappa^{-3}) & t = 3. \end{cases} \tag{B.2}$$

noting $\gamma(y)$ is a symmetric function, from (B.2), the conditional expectation γ_i is given by

$$\begin{aligned}\gamma_i &= \int_{-\pi}^{\pi} \gamma(\Theta_i - \theta_j) f(\theta_j) d\theta_j \\ &= \int_{-\pi}^{\pi} \gamma(y) f(\Theta_i + y) dy \\ &= \sum_{t=0}^{2} \frac{f^{(2t)}(\Theta_i)}{(2t)!} \int_{-\pi}^{\pi} \gamma(y) y^{2t} dy + O\left(\int \gamma(y) y^6 dy\right) \\ &= -f(\Theta_i) + f^{(4)}(\Theta_i)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}).\end{aligned}$$

Appendix C

We derive each term of the variance $\mathrm{Var}_f[\overline{\mathrm{CV}}(\kappa)]$. We present the expectation $\mathrm{E}_f[\gamma_{ij}^2]$ as

$$\begin{aligned}\mathrm{E}_f[\gamma_{ij}^2] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \gamma^2(\theta_i - \theta_j) f(\theta_i) f(\theta_j) d\theta_i d\theta_j \\ &= \int_{-\pi}^{\pi} f(\theta_j) \int_{-\pi}^{\pi} \gamma^2(u) f(\theta_j + u) du d\theta_j \\ &= \int f(\theta_j) \int_{-\pi}^{\pi} \gamma^2(u)[f(\theta_j) + O(u^2)] du d\theta_j \\ &= R(f)R(\gamma) + O(R(\gamma(y)y)) \end{aligned} \tag{C.1}$$

We produce the following lemma regarding $R(\gamma(y)y^t)$.

Lemma C. 1. We set $Q_{2t}(L) := \int_{-\infty}^{\infty}\left\{2^{-1}\mu_0^{-2}(L)\eta(z) - 2^{1/2}\mu_0^{-1}(L)L(z^2/2)\right\}^2 z^{2t}dz$. Then, the term $R(\gamma(y)y^t)$ is given by

$$R(\gamma(y)y)^t = \kappa^{-(2t-1)/2}[Q_{2t}(L) + o(1)] \quad t = 0, 1.$$

Proof. Let $y = \kappa^{-1/2}z$. Then, Applying $\cos(\kappa^{-1/2}z) = 1 - z^2/(2\kappa) + O(\kappa^{-2})$, the Taylor expansion of $L_\kappa(\kappa^{-1/2}z)$ is given by

$$\begin{aligned}L_\kappa(\kappa^{-1/2}z) &= L(\kappa[1 - \{1 - z^2/(2\kappa) + O(\kappa^{-2})\}]) \\ &= L(z^2/2) + O(\kappa^{-1}).\end{aligned} \tag{C.2}$$

It follows from (C.2) that

$$\begin{aligned}\int_{-\pi}^{\pi} L_\kappa(w)L_\kappa(w + \kappa^{-1/2}z)dw &= \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L_\kappa(\kappa^{-1/2}t)L_\kappa(\kappa^{-1/2}(t+z))\kappa^{-1/2}dt \\ &= \kappa^{-1/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L(t^2/2)L((t+z)^2/2)dt + O(\kappa^{-3/2}) \\ &= \kappa^{-1/2}[\eta(z) + o(1)].\end{aligned} \tag{C.3}$$

We put $Q_{\kappa^{1/2},2t}(L) := \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\left\{2^{-1}\mu_0^{-2}(L)\eta(z) - 2^{1/2}\mu_0^{-1}(L)L(z^2/2)\right\}^2 z^{2t}dz$. Then it holds from (b) and (e) that $Q_{\kappa^{1/2},2t}(L) = Q_{2t}(L) + o(1)$ for $t = 0, 1$. By combining (C.3) and Lemma A.1, the term $R(\gamma(y)y^t)$ is given by

$$
\begin{aligned}
R(\gamma(y)y^t) &= \int_{-\pi}^{\pi}\left\{\int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w+y)dw - 2K_\kappa(y)\right\}^2 y^{2t}dy \\
&= \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\left\{\int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w+\kappa^{-1/2}z)dw - 2K_\kappa(\kappa^{-1/2}z)\right\}^2 (\kappa^{-1/2}z)^{2t}\kappa^{-1/2}dz \\
&= \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\left\{C_\kappa^{-2}(L)\int_{-\pi}^{\pi} L_\kappa(w)L_\kappa(w+\kappa^{-1/2}z)dw - 2K_\kappa(\kappa^{-1/2}z)\right\}^2 (\kappa^{-1/2}z)^{2t}\kappa^{-1/2}dz \\
&= \kappa^{-(2t+1)/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\left\{C_\kappa^{-2}(L)\kappa^{-1/2}[\eta(z)+o(1)] - 2C_\kappa^{-1}(L)[L(z^2/2)+O(\kappa^{-1})]\right\}^2 z^{2t}dz \\
&= \kappa^{-(2t+1)/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\Big\{(\kappa^{-1/2}2^{1/2}\mu_0(L)+O(\kappa^{-3/2}))^{-2}\kappa^{-1/2}[\eta(z)+o(1)] \\
&\quad - 2(\kappa^{-1/2}2^{1/2}\mu_0(L)+O(\kappa^{-3/2}))^{-1}[L(z^2/2)+O(\kappa^{-1})]\Big\}^2 z^{2t}dz \\
&= \kappa^{-(2t+1)/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi}\left[\kappa^{1/2}\left\{2^{-1}\mu_0^{-2}(L)\eta(z) - 2^{1/2}\mu_0^{-1}(L)L(z^2/2)+o(1)\right\}\right]^2 z^{2t}dz \\
&= \kappa^{-(2t-1)/2}\left[Q_{\kappa,2t}(L)+o(1)\right] \\
&= \kappa^{-(2t-1)/2}[Q_{2t}(L)+o(1)].
\end{aligned}
$$

□ □

Noting that $Q_0(L) = Q(L)$, from combining (C.1) and Lemma C.1, the expectation $\mathrm{E}_f[\gamma_{ij}^2]$ is given by

$$
\mathrm{E}_f[\gamma_{ij}^2] = \kappa^{1/2}[Q(L)R(f)+o(1)]. \tag{C.4}
$$

From combining (3.7) and (C.4), it follows that $\mathrm{Var}[\gamma_{ij}]$ is equivalent to (3.9).

Noting that $\gamma_i = \int_{-\pi}^{\pi}\gamma(\theta_i-\theta_j)f(\theta_j)d\theta_j$, then, from (3.6) we derive $\mathrm{E}_f[\gamma_{ij}\gamma_{ik}]$. That is,

$$
\begin{aligned}
\mathrm{E}_f[\gamma_{ij}\gamma_{ik}] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\gamma(\theta_i-\theta_j)\gamma(\theta_i-\theta_k)f(\theta_i)f(\theta_j)f(\theta_k)d\theta_i d\theta_j d\theta_k \\
&= \int_{-\pi}^{\pi} f(\theta_i)\left[\int_{-\pi}^{\pi}\gamma(\theta_i-\theta_j)f(\theta_j)d\theta_j\right]^2 d\theta_i \\
&= R(f^{3/2}) - 2R((f^{(4)})^{1/2}f)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}).
\end{aligned} \tag{C.5}
$$

By combining (3.7) and (C.5), we obtain that $\mathrm{Cov}_f[\gamma_{ij},\gamma_{ik}]$ is equivalent to (3.11).

From (3.7), we derive that $\mathrm{E}_f[\gamma_{ij}f(\Theta_i)]$ is given by

$$
\begin{aligned}
\mathrm{E}_f[\gamma_{ij}f(\Theta_i)] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\gamma(\theta_i-\theta_j)f(\theta_i)f(\theta_j)f(\theta_i)d\theta_i d\theta_j \\
&= -R(f^{3/2}) + R((f^{(4)})^{1/2}f)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}).
\end{aligned} \tag{C.6}
$$

From combining (3.7) and (C.6), we derive that $\mathrm{Cov}_f[\gamma_{ij}, f(\Theta_i)]$ is given by (3.12).

The variance $\mathrm{Var}_f[f(\Theta_i)]$ is equivalent to

$$\begin{aligned}\mathrm{Var}_f[f(\Theta_i)] &= \mathrm{E}[f^2(\Theta_i)] - \mathrm{E}[f(\Theta_i)]^2 \\ &= R(f^{3/2}) - R(f)^2 \\ &= I_3.\end{aligned}$$

Appendix D

we derive the expectation $\mathrm{E}_f[U_{ij}^{2m}]$. That is,

$$\begin{aligned}\mathrm{E}_f[U_{ij}^{2m}] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} T_g^{(4)}(\theta_i-\theta_j)^{2m} f(\theta_i)f(\theta_j)d\theta_i d\theta_j \\ &= \int_{-\pi}^{\pi} f(\theta_j)\int_{-\pi}^{\pi} T_g^{(4)}(u)^{2m} f(\theta_j+u)dud\theta_j \\ &= \int_{-\pi}^{\pi} f(\theta_j)\int_{-\pi}^{\pi} T_g^{(4)}(u)^{2m}[f(\theta_j)+O(u^2)]dud\theta_j \\ &= \psi_0 R(\{T^{(4)}\}_g^{2m}) + O(R(\{T_g^{(4)}(u)\}^{2m}u) \qquad \text{(D.1)}\end{aligned}$$

Lemma D. 1. The term $R(\{T_g^{(4)}(\theta)\}^m\theta^t)$ is given by

$$R(\{T_g^{(4)}(\theta)\}^m\theta^t) = g^{(10m-2t-1)/2}\left\{G_{m,t}(S_4) + o(1)\right\}, \qquad \text{(D.2)}$$

for $t = 0, 1$ and $m = 0, 1$.

Proof. The Taylor expansions of $\cos(g^{-1/2}z)$ and $\sin g^{-1/2}z$ are reduced to

$$\cos(g^{-1/2}z) = 1 - z^2/(2g) + O(g^{-2}), \qquad \text{(D.3)}$$

and,

$$\sin(g^{-1/2}z) = g^{-1/2}z + O(g^{-3/2}), \qquad \text{(D.4)}$$

respectively. From considering (3.20), (C.2), (D.3), and (D.4), the approximation of $S_g^{(4)}(g^{-1/2}z)$ is given by

$$\begin{aligned}S_g^{(4)}(g^{-1/2}z) &= g^2\{S^{(2)}(z^2/2) + 6z^2S^{(3)}(z^2/2) + z^4S^{(4)}(z^2/2) + o(1)\} \\ &= g^2\{S_4(z^2/2) + o(1)\}. \qquad \text{(D.5)}\end{aligned}$$

We set $\delta_{g^{1/2},t}(S_4^m) := \int_{-g^{1/2}\pi}^{g^{1/2}\pi} S_4^{2m}(z^2/2)z^{2t}dz$. Then, it holds from (g) that $\delta_{g^{1/2},t}(S_4^m) = \delta_t(S_4^m) + o(1)$ for $t = 0, 1$, and $m = 1, 2$. By combining Lemma A.3 and (D.5), The term $R(\{T_g^{(4)}(\theta)\}^m\theta^t)$ is reduced to

$$\begin{aligned}
R(\{T_g^{(4)}(\theta)\}^m\theta^t) &= C_g^{-2m}(S)\int_{-\pi}^{\pi}\left\{S_g^{(4)}(\theta)^m\theta^t\right\}^2 d\theta \\
&= C_g^{-2m}(S)\int_{-g^{1/2}\pi}^{g^{1/2}\pi}\left\{S_g^{(4)}(g^{-1/2}z)^m(g^{-1/2}z)^t\right\}^2 g^{-1/2}dz \\
&= C_g^{-2m}(S)g^{-(2t+1)/2}\int_{-g^{1/2}\pi}^{g^{1/2}\pi}[g^2\{S_4(z^2/2)+o(1)\}]^{2m}z^{2t}dz \\
&= \{2^{1/2}\mu_0^{-1}(S)g^{-1/2}+O(g^{-(p+1)/2})\}^{-2m}g^{(8m-2t+1)/2}\left\{\delta_{g^{1/2},t}(S_4^m)+o(1)\right\} \\
&= 2^{-m}\mu_0^{2m}(S)g^{(10m-2t-1)/2}\left\{\delta_t(S_4^m)+o(1)\right\} \\
&= g^{(10m-2t-1)/2}\left\{G_{m,t}(S_4)+o(1)\right\}.
\end{aligned}$$

□ □

From combining (D.1), and Lemma D.1, the expectation $\mathrm{E}_f[U_{ij}^{2m}]$ is given by

$$\mathrm{E}_f[U_{ij}^{2m}] = g^{(10m-1)/2}[\psi_0 G_{m,0}(S_4)+o(1)]. \tag{D.6}$$

It follows from (3.26) that

$$\begin{aligned}
\mathrm{E}_f[U_i^2] &= \mathrm{E}_f[\{f^{(4)}(\Theta_i)+o(1)\}^2] \\
&= \mathrm{E}_f[f^{(4)}(\Theta_i)^2]+o(1).
\end{aligned}$$

Appendix E

We calculate $\frac{d}{d\kappa}\gamma(y_{ij})$. We derive

$$\frac{d}{d\kappa}L_\kappa(w)L_\kappa(w+y) = L_\kappa'(w)L_\kappa(w+y)\{1-\cos(w)\}+L_\kappa(w)L_\kappa'(w+y)\{1-\cos(w+y)\}. \tag{E.1}$$

We set $\frac{d}{d\kappa}C_\kappa(L)=C_\kappa'(L)$ and $\alpha_t(\phi_\kappa) := \int_{-\pi}^{\pi}\phi_\kappa(y)y^t dy$. It follows that

$$\begin{aligned}
\kappa C_\kappa^{-1}(L)C_\kappa'(L) &= \kappa C_\kappa^{-1}(L)\int_{-\pi}^{\pi}\frac{d}{d\kappa}L_\kappa(\theta)d\theta \\
&= \alpha_0(\phi_\kappa)
\end{aligned} \tag{E.2}$$

We provide the following lemma regarding $\alpha_t(\phi_\kappa)$

Lemma E. 1. The term $\alpha_\kappa(\phi_\kappa)$ is given by

$$\alpha_t(\phi_\kappa) = \begin{cases} -\frac{1}{2}-\frac{3}{8}\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}+O(\kappa^{-2}) & t=0, \\ -3\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}+O(\kappa^{-2}) & t=2, \\ 24\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2}=O(\kappa^{-2}) & t=4. \end{cases}$$

Proof. From (b), the partial integration of $\mu_{\kappa,l}(L') := \int_0^\kappa L(r)r^{(l-1)/2}dr$ for $l \leq 4$ is to

$$\begin{aligned}\mu_{\kappa,l}(L') &= [L(r)r^{(l-1)/2}]_0^\kappa - \frac{l-1}{2}\int_0^\kappa L(r)r^{(l-3)/2}dr \\ &= -\frac{l-1}{2}\mu_{\kappa,l-2}(L) + O(\kappa^{-3}). \end{aligned} \tag{E.3}$$

The term $\alpha_{2t}(\phi_\kappa)$ is divided into the following two terms. That is,

$$\alpha_{2t}(\phi_\kappa) = 2\int_0^{\pi/2}\phi_\kappa(\theta)\theta^{2t}d\theta + 2\int_{\pi/2}^{\pi}\phi_\kappa(\theta)\theta^{2t}d\theta. \tag{E.4}$$

Recalling that we chose the second-order kernel for LSCV, the second term of (E.4) is ignored from combining (d), (E.3), and lemma A.1. That is,

$$\begin{aligned}2\int_{\pi/2}^{\pi}\phi_\kappa(\theta)\theta^{2t}d\theta &\leq 2\pi^{2t}\int_{\pi/2}^{\pi}\phi_\kappa(\theta)d\theta \\ &\leq 2\pi^{2t}C_\kappa^{-1}(L)\int_{\pi/2}^{\pi}L'(\kappa\{1-\cos(\theta)\})\kappa\{1-\cos(\theta)\}d\theta \\ &= 2\pi^{2t}C_\kappa^{-1}(L)\int_\kappa^{2\kappa}L'(r)r\{r\kappa(2-r/\kappa)\}^{-1/2}dr \\ &= 2\pi^{2t}C_\kappa^{-1}(L)\kappa^{-1/2}\int_\kappa^{2\kappa}L'(r)r^{1/2}dr\{2^{-1/2}+O(\kappa^{-1})\} \\ &= O(\kappa^{-3}). \end{aligned} \tag{E.5}$$

By considering (d), (E.3), and (E.4),we derive the terms $\alpha_0(\phi_\kappa)$, $\alpha_2(\phi_\kappa)$, and $\alpha_4(\phi_\kappa)$. That is,

$$\begin{aligned}\alpha_0(\phi_\kappa) &= 2\int_0^{\pi/2}\phi_\kappa(\theta)d\theta + O(\kappa^{-3}) \\ &= 2C_\kappa^{-1}(L)\int_0^\kappa L'(r)r\{r\kappa(2-r\kappa)\}^{-1/2}dr + O(\kappa^{-3}) \\ &= 2C_\kappa^{-1}(L)\kappa^{-1/2}\int_0^\kappa L'(r)r^{1/2}[2^{-1/2} - 2^{-5/2}r/\kappa + O(\kappa^{-2})]dr + O(\kappa^{-3}) \\ &= 2C_\kappa^{-1}(L)\kappa^{-1/2}[2^{-1/2}\mu_{2,\kappa}(L') - 2^{-5/2}\kappa^{-1}\mu_{4,\kappa}(L') + O(\kappa^{-2})] + O(\kappa^{-3}) \\ &= -\frac{1}{2} - \frac{3}{8}\mu_0^{-1}(L)\mu_2(L)\kappa^{-1} + O(\kappa^{-2}), \end{aligned} \tag{E.6}$$

$$\begin{aligned}\alpha_2(\phi_\kappa) &= 2C_\kappa^{-1}(L)\int_0^{\pi/2}L'(\kappa\{1-\cos(\theta)\})\kappa\{1-\cos(\theta)\}\theta^2d\theta + O(\kappa^{-3}) \\ &= 2C_\kappa^{-1}(L)\int_0^\kappa L'(r)r[r/\kappa(2-r/\kappa)+O(\kappa^{-2})]\{r\kappa(2-r/\kappa)\}^{-1/2}dr + O(\kappa^{-3}) \\ &= 2C_\kappa^{-1}(L)\kappa^{-3/2}\int_0^\kappa L'(r)r^{3/2}(2-r/\kappa)^{1/2}dr + O(\kappa^{-2}) \\ &= 2C_\kappa^{-1}(L)\kappa^{-3/2}\mu_{4,\kappa}(L')\{2^{1/2}+O(\kappa^{-1})\} + O(\kappa^{-2}) \\ &= 2\mu_0^{-1}(L)(-3\mu_2(L)/2)\kappa^{-1} + O(\kappa^{-2}) \\ &= -3\mu_0^{-1}(L)\mu_2(L)\kappa^{-1} + O(\kappa^{-2}), \end{aligned}$$

and,

$$\begin{aligned}\alpha_4(\phi_\kappa) &= 2\int_0^{\pi/2}\phi_\kappa(\theta)\theta^4 d\theta + O(\kappa^{-3})\\ &= 2C_\kappa^{-1}(L)\int_0^\kappa L'(r)r[\{r/\kappa(2-r/\kappa)\}^2 + O(\kappa^{-3})]\{r\kappa(2-r/\kappa)\}^{-1/2}dr + O(\kappa^{-3})\\ &= O(\kappa^{-2}). \end{aligned} \tag{E.7}$$

□ □

Then, by combining (E.1), (E.2), and Lemma E.1, it follows that

$$\begin{aligned}\frac{d\gamma(y_{ij})}{d\kappa} &= \frac{d}{d\kappa}\left\{C_\kappa^{-2}(L)\int_{-\pi}^{\pi}L_\kappa(w)L_\kappa(w+y_{ij})dw - 2C_\kappa^{-1}(L)L_\kappa(y_{ij})\right\}\\ &= -2C_\kappa^{-3}(L)C'_\kappa(L)\int_{-\pi}^{\pi}L_\kappa(w)L_\kappa(w+y_{ij})dw\\ &\quad + C_\kappa^{-2}(L)\int_{-\pi}^{\pi}\frac{d}{d\kappa}\{L_\kappa(w)L_\kappa(w+y_{ij})\}dw\\ &\quad + 2C_\kappa^{-2}(L)C'_\kappa(L)L_\kappa(y_{ij}) - 2C_\kappa^{-1}(L)\frac{d}{d\kappa}L_\kappa(y_{ij})\\ &= \kappa^{-1}\Bigg[-2\alpha_0(\phi_\kappa)\int_{-\pi}^{\pi}K_\kappa(w)K_\kappa(w+y_{ij})dw\\ &\quad + \int_{-\pi}^{\pi}\{\phi_\kappa(w)K_\kappa(w+y_{ij}) + K_\kappa(w)\phi_\kappa(w+y_{ij})\}dw\\ &\quad + 2\alpha_0(\phi_\kappa)K_\kappa(y_{ij}) - 2\phi_\kappa(y_{ij})\Bigg]\\ &= \kappa^{-1}\Bigg[\int_{-\pi}^{\pi}K_\kappa(w)K_\kappa(w+y_{ij})dw - 2K_\kappa(y_{ij})\\ &\quad + K_\kappa(y_{ij}) + \int_{-\pi}^{\pi}\{\phi_\kappa(w)K_\kappa(w+y_{ij}) + K_\kappa(w)\phi_\kappa(w+y_{ij})\}dw - 2\phi_\kappa(y_{ij})\\ &\quad + \frac{3}{4}\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\left\{\int_{-\pi}^{\pi}K_\kappa(w)K_\kappa(w+y_{ij})dw - K_\kappa(y_{ij})\right\}\Bigg]\\ &= \kappa^{-1}[\gamma(y_{ij}) + \rho(y_{ij}) + 3/4\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\tau(y_{ij})]\\ &= \kappa^{-1/2}V_{ij}. \end{aligned} \tag{E.8}$$

We obtain the equation (4.5) from (E.8) and (4.4).

Appendix F

Let $\rho_i := \mathrm{E}_f[\rho_{ij}|\Theta_i]$ and $\tau_i := \mathrm{E}_f[\tau_{ij}|\Theta_i]$. Then, The conditional expectation V_i is presented as the following linear combination of the conditional expectations γ_i, ρ_i, and τ_i.

$$V_i = \kappa^{-1/2}\left[\gamma_i + \rho_i + \frac{3}{4}\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\tau_i\right]. \tag{F.1}$$

We present the following lemma regarding ρ_i.

Lemma F. 1. The conditional expectation ρ_i is given by

$$\rho_i = f(\Theta_i) - 3\left[\frac{f^{(2)}(\Theta_i)}{4} + f^{(4)}(\Theta_i)\right]\mu_0^{-2}\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}) \tag{F.2}$$

Proof. The term $\alpha_{2t}(\rho) = \int_{-\pi}^{\pi}\rho(y)y^{2t}dy$ is given by

$$\begin{aligned}\alpha_{2t}(\rho) &= \alpha_{2t}(K_\kappa)\\ &\quad + \kappa C_\kappa^{-2}(L)\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\{\phi_\kappa(w)K_\kappa(w+y) + K_\kappa(w)\phi_\kappa(w+y)\}y^{2t}dwdy - 2\alpha_{2t}(\phi_\kappa).\end{aligned} \tag{F.3}$$

The second term of (F.3) is reduced to

$$\begin{aligned}&\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\{\phi_\kappa(w)K_\kappa(w+y) + K_\kappa(w)\phi_\kappa(w+y)\}y^{2t}dwdy\\ &\quad = 2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}K_\kappa(w)\phi_\kappa(s)\}(s-w)^{2t}dwdy\\ &\quad = 2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}K_\kappa(w)\phi_\kappa(s)\left[\sum_{m=0}^{2t}(-1)^m{}_{2t}C_m w^m s^{2t-m}\right]dwds\\ &\quad = 2\sum_{m=0}^{2t}(-1)^m{}_{2t}C_m\alpha_m(K_\kappa)\alpha_{2t-m}(\phi_\kappa).\end{aligned} \tag{F.4}$$

It follows from (F.3) and (F.4) that

$$\alpha_0(\rho) = \alpha_0(K_\kappa) = 1, \tag{F.5}$$

and

$$\alpha_{2t}(\rho) = \alpha_{2t}(K_\kappa) + 2\sum_{m=1}^{2t}(-1)^m{}_{2t}C_m\alpha_m(K_\kappa)\alpha_{2t-m}(\phi_\kappa) \quad t \geq 1. \tag{F.6}$$

From combining Lemma A.2, Lemma E.1, and (F.6), it follows that

$$\alpha_2(\rho) = -\frac{3}{2}\mu_0^{-2}\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}), \tag{F.7}$$

$$\alpha_4(\rho) = -72\mu_0^{-2}\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}), \tag{F.8}$$

and,

$$\alpha_6(\rho) = O(\kappa^{-3}). \tag{F.9}$$

By combining (F.5), (F.7), (F.8) and (F.9), we obtain the conditional expectation ρ_i. That is,

$$\begin{aligned}
\rho_i &= \int_{-\pi}^{\pi} \rho(\theta_j - \Theta_i) f(\theta_j) d\theta_j \\
&= \int_{-\pi}^{\pi} \rho(y) f(\Theta_i + y) dy \\
&= \sum_{t=0}^{2} \frac{f^{(2t)}(\Theta_i)}{(2t)!} \alpha_{2t}(\rho) + O(\alpha_6(\rho)) \\
&= f(\Theta_i) - 3\left[\frac{f^{(2)}(\Theta_i)}{4} + f^{(4)}(\Theta_i)\right] \mu_0^{-2} \mu_2^2(L) \kappa^{-2} + O(\kappa^{-3}). \qquad \text{(F.10)}
\end{aligned}$$

□ □

We present the following lemma regarding τ_i.

Lemma F. 2. The conditional expectation τ_i is given by

$$\tau_i = f^{(2)}(\Theta_i) \mu_0^{-1} \mu_2(L) \kappa^{-1} + O(\kappa^{-2}). \qquad \text{(F.11)}$$

Proof. We set $\alpha_t(\tau) := \int_{-\pi}^{\pi} \tau(y) y^t dy$. Then, it follows that

$$\alpha_{2t}(\tau) = \sum_{m=0}^{2t} (-1)^m {}_{2t}C_m \alpha_m(K_\kappa) \alpha_{2t-m}(K_\kappa) - \alpha_{2t}(K_\kappa). \qquad \text{(F.12)}$$

From combining Lemma A. 2 and (F.12) It follows that the terms $\alpha_0(\tau)$, $\alpha_2(\tau)$ and $\alpha_4(\tau)$ are equal to,

$$\alpha_0(\tau) = 0, \qquad \text{(F.13)}$$

$$\alpha_2(\tau) = 2\mu_0^{-1}(L) \mu_2(L) \kappa^{-1} + O(\kappa^{-1}), \qquad \text{(F.14)}$$

and,

$$\alpha_4(\tau) = O(\kappa^{-2}), \qquad \text{(F.15)}$$

respectively. It is shown from (F.13), (F.14) and (F.15) that

$$\begin{aligned}
\tau_i &= \int_{-\pi}^{\pi} \tau(\theta_j - \Theta_i) f(\theta_j) d\theta_j \\
&= \int_{-\pi}^{\pi} \tau(y) f(\Theta_i + y) dy \\
&= f(\Theta_i) \alpha_0(\tau) + \frac{f^{(2)}(\Theta_i)}{2} \alpha_2(\tau) + O(\alpha_4(\tau)) \\
&= f^{(2)}(\Theta_i) \mu_0^{-1}(L) \mu_2(L) \kappa^{-1} + O(\kappa^{-2}).
\end{aligned}$$

□

By combining (F.1),(3.6) Lemma F.1, and Lemma F.2, The conditional expectation V_i is reduced to

$$\begin{aligned} V_i &= \kappa^{-1/2}\left[\gamma_i + \rho_i + \frac{3}{4}\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\tau_i\right] \\ &= -2f^{(4)}(\Theta_i)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-5/2} + o(\kappa^{-5/2}). \end{aligned} \tag{F.16}$$

The expectations of V_i and V_i^2 are given by

$$\mathrm{E}_f[V_i] = -2R(f'')\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-5/2} + o(\kappa^{-5/2}), \tag{F.17}$$

and

$$\mathrm{E}_f[V_i^2] = 4[R(f^{(4)}f^{1/2})]\mu_0^{-4}(L)\mu_2^4(L)\kappa^{-5} + o(\kappa^{-5}), \tag{F.18}$$

respectively. We obtain the variance of X_i from (F.17) and (F.18). That is,

$$\mathrm{Var}_f[X_i] = 4[R(f^{(4)}f^{1/2}) - R(f'')^2]\mu_0^{-4}(L)\mu_2^4(L)\kappa^{-5} + o(\kappa^{-5}). \tag{F.19}$$

From (F.19), we show that the variance $\mathrm{Var}_f[X_i]$ is finite. Thus, we obtain (4.7) from the central limit theorem.

Appendix G

We derive the expectation $\mathrm{E}_f[V_{ij}^{2m}]$. That is,

$$\begin{aligned} \mathrm{E}_f[V_{ij}^{2m}] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\kappa^{-1/2}\{\gamma(\theta_i-\theta_j) + \rho(\theta_i-\theta_j) + O(\kappa^{-1})\}]^{2m} f(\theta_i)f(\theta_j)d\theta_i d\theta_j \\ &= \kappa^{-m}\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\{\gamma(\theta_i-\theta_j) + \rho(\theta_i-\theta_j)\}^{2m} f(\theta_i)f(\theta_j)d\theta_i d\theta_j\right]\{1+o(1)\} \\ &= \kappa^{-m}\left[R((\gamma+\rho)^m)R(f) + O(R((\gamma+\rho)^m y))\right]. \end{aligned} \tag{G.1}$$

Lemma G. 1. The term $R((\gamma+\rho)^m y)^m y)$ is given by

$$R((\gamma+\rho)^m)^m y) = \kappa^{(2m-2t-1)/2}[M_{m,t}(L) + o(1)]$$

Proof. We set

$$\psi(y) = \int_{-\pi}^{\pi} L'(\kappa\{1-\cos(w)\})\kappa\{1-\cos(w)\}L(\kappa\{1-\cos(w+y)\})dw.$$

then, the term $\int_{-\pi}^{\pi}\frac{d}{d\kappa}\{L_\kappa(w)L_\kappa(w+y)\}dw$ is reduced to

$$\kappa\int_{-\pi}^{\pi}\frac{d}{d\kappa}\{L_\kappa(w)L_\kappa(w+y)\}dw = \psi(y) + \psi(-y). \tag{G.2}$$

We set $\lambda_{\kappa^{1/2}}(L) := \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L'(t^2/2)L((t+z)^2/2)t^2/2dt$. Then, it holds that $\lambda_{\kappa^{1/2}}(L) = \lambda(L) + o(1)$ from (f). Thus, it follows that

$$\begin{aligned}
\psi(\kappa^{-1/2}z) &= \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L'(\kappa\{1-\cos(\kappa^{-1/2}t)\})\kappa\{1-\cos(\kappa^{-1/2}t)\}L(\kappa\{1-\cos(\kappa^{-1/2}(t+z))\})\kappa^{-1/2}dt \\
&= \kappa^{-1/2}\left[\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L((t+z)^2/2)L'(t^2/2)t^2/2dt + O(\kappa^{-1})\right] \\
&= \kappa^{-1/2}\left[\lambda_{\kappa^{1/2}}(z) + O(\kappa^{-1})\right] \\
&= \kappa^{-1/2}\left[\{\lambda(z) + o(1)\} + O(\kappa^{-1})\right] \\
&= \kappa^{-1/2}[\lambda(z) + o(1)] \qquad \text{(G.3)}
\end{aligned}$$

We set $M_{\kappa,m,t}(L) := \int_{-\kappa\pi}^{\kappa\pi} m(L)^{2m} z^{2t} dz$. Then, it holds from combining (b), (e), and (f) that $M_{\kappa,m,t}(L) = M_{m,t}(L) + o(1)$. From considering this, (C.3), and (G.3), The term $R(\{\gamma+\rho\}^m y^t)$ is reduced to

$$\begin{aligned}
R((\gamma+\rho)^m y^t) &= \int_{-\pi}^{\pi} \{\gamma(y) + \rho(y)\}^{2m} y^{2t} dy \\
&= \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} \Bigg[C_\kappa^{-2}(L)\left\{\int_{-\pi}^{\pi} L_\kappa(w)L_\kappa(w+\kappa^{-1/2}z)dw + \psi(\kappa^{-1/2}z) + \psi(-\kappa^{-1/2}z)\right\} \\
&\quad - C_\kappa^{-1}(L)\{L_\kappa(\kappa^{-1/2}z) + 2L'(\kappa\{1-\cos(\kappa^{-1/2}z)\})\kappa\{1-\cos(\kappa^{-1/2}z)\}\}\Bigg]^{2m} (\kappa^{-1/2}z)^{2t}\kappa^{-1/2}dz \\
&= \kappa^{-(2t+1)/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} \Bigg[C_\kappa^{-2}(L)\kappa^{-1/2}\{\eta(z) + \lambda(z) + \lambda(-z) + o(1)\} \\
&\quad - C_\kappa^{-1}(L)\{L(z^2/2) + L'(z^2/2)z^2 + O(\kappa^{-1})\}\Bigg]^{2m} z^{2t}dz \\
&= \kappa^{-(2t+1)/2}\int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} \left[\kappa^{1/2}\left\{\frac{\eta(z)+\lambda(z)+\lambda(-z)}{2\mu_0^2(L)} - \frac{L(z^2/2)+L'(z^2/2)z^2}{2^{1/2}\mu_0(L)} + o(1)\right\}\right]^{2m} z^{2t}dz \\
&= \kappa^{(2m-2t-1)/2}[M_{\kappa,m,t}(L) + o(1)] \\
&= \kappa^{(2m-2t-1)/2}[M_{m,t}(L) + o(1)].
\end{aligned}$$

□ □

From combining (G.4), and Lemma G.1, it follows that

$$\mathrm{E}_f[V_{ij}^{2m}] = \kappa^{-1/2}[M_{m,0}(L)R(f) + o(1)]. \qquad \text{(G.4)}$$

From (F.16), it follows that $V_i = O(\kappa^{-5/2})$. Then, The expectation $\mathrm{E}_f[H_{ij}^{2m}]$ is reduced to

$$\begin{aligned}
\mathrm{E}_f[H_{ij}^{2m}] &= \mathrm{E}_f[\{V_{ij} - V_i - V_j + \mathrm{E}_f[V_{ij}]\}^{2m}] \\
&= \mathrm{E}_f[V_{ij}^{2m}]\{1 + o(1)\} \\
&= \kappa^{-1/2}[M_{m,0}(L)R(f) + o(1)]. \qquad \text{(G.5)}
\end{aligned}$$

Noting that V_{ii} is a constant, it follows that

$$\begin{aligned}
G_{ij} &:= \mathrm{E}[H_{ii}H_{ij}] \\
&= \mathrm{E}[\{V_{ii} + \mathrm{E}_f[V_i] - 2V_i\}\{V_{ij} - V_i - V_j + \mathrm{E}_f[V_i]\}] \\
&= 0 - 2\mathrm{E}_f[V_i^2] + 2\mathrm{E}_f[V_i^2] + 2\mathrm{E}_f[V_i]^2 - 2\mathrm{E}_f[V_i]^2 \\
&= 0. \qquad \text{(G.6)}
\end{aligned}$$

From (G.5) and (G.6), it follows that the U-statistic H_{ij} satisfies (4.3). That is,

$$\begin{aligned}
\frac{\mathrm{E}[G_{ij}^2] + n^{-1}\mathrm{E}_f[H_{ij}^4]}{\mathrm{E}_f[H_{ij}^2]^2} &= \frac{n^{-1}[\kappa^{-1/2}[M_{2,0}(L)R(f) + o(1)]}{[\kappa^{-1/2}M_{1,0}(L)R(f) + o(1)]^2} \\
&= o(1). \qquad \text{(G.7)}
\end{aligned}$$

We obtain the asymptotic normality for (4.8) from (G.7).

Appendix-H

Let $g^{-9/4}W_{ij} = Q_{ij}$. By (3.31), the expectation $\mathrm{E}_f[Q_{ij}^2]$ is given by

$$\begin{aligned}
\mathrm{E}_f[Q_{ij}^2] &= g^{-9/2}\mathrm{E}_f[W_{ij}^2] \\
&= G_{1,0}(S_4)\psi_0 + o(1). \qquad \text{(H.1)}
\end{aligned}$$

From combining (3.26), (3.27) and Lemma D.1, it follows that

$$\begin{aligned}
\mathrm{E}_f[Q_{ij}^4] &= g^{-9}\mathrm{E}_f[W_{ij}^4] \\
&= g^{-9}\mathrm{E}_f[U_{ij}^4]\{1 + o(1)\} \\
&= g^{1/2}\{G_{2,0}(S_4)\psi_0 + o(1)\}. \qquad \text{(H.2)}
\end{aligned}$$

By combining $G_{ij} = 0$, (H.1), and (H.2) It follows that that

$$\begin{aligned}
\frac{\mathrm{E}[G_{ij}^2] + n^{-1}\mathrm{E}_f[Q_{ij}^4]}{\mathrm{E}_f[Q_{ij}^2]^2} &= \frac{0 + n^{-1}[G_{2,0}(S_4)\psi_0 g^{1/2} + o(g^{1/2})]}{[G_{1,0}(S_4)\psi_0 + o(1)]^2} \\
&= o(1). \qquad \text{(H.3)}
\end{aligned}$$

the d-generate U statistic Q_{ij} satisfies Lemma 1 by (H.3). Therefore, as $n \to \infty$, it holds that

$$\sum_{i<j} Q_{ij} \xrightarrow{d} N(0, n^2 G_{1,0}(S_4)\psi_0/2). \qquad \text{(H.4)}$$

We obtain the asymptotic normality from (4.17) from (H.4).