九州大学学術情報リポジトリ
Kyushu University Institutional Repository

ASYMPTOTIC PROPERTY OF WRAPPED CAUCHY KERNEL DENSITY ESTIMATION ON THE CIRCLE

Tsuruta, Yasuhito
Graguate School of Human and Socio-Environment Studies, Kanazawa University

Sagae, Masahiko
School of Economics, Kanazawa University

https://doi.org/10.5109/2232318

出版情報：Bulletin of informatics and cybernetics. 49, pp.1-10, 2017-12. Research Association of Statistical Sciences
バージョン：
権利関係：

KYUSHU UNIVERSITY

ASYMPTOTIC PROPERTY OF WRAPPED CAUCHY KERNEL DENSITY ESTIMATION ON THE CIRCLE

by

Yasuhito Tsuruta and **Masahiko Sagae**

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.49

FUKUOKA, JAPAN
2017

Bulletin of Informatics and Cybernetics, Vol. 49, 2017

ASYMPTOTIC PROPERTY OF WRAPPED CAUCHY KERNEL DENSITY ESTIMATION ON THE CIRCLE

By

Yasuhito Tsuruta[*] and **Masahiko Sagae**[†]

Abstract

We discuss theoretical properties of the Wrapped Cauchy (WC) kernel. In this paper, we show that the WC kernel has the convergent rate of $O(n^{-2/3})$ and holds the asymptotic normality. The rate of the WC kernel is not better than the rate $O(n^{-4/5})$ of the von Mises (VM) kernel. However, some numerical experiments show the better behavior of the WC kernel rather than the VM kernel under the condition of the multimodality and/or the heavy tail.

Key Words and Phrases: Kernel density estimation, Circular data, Wrapped Cauchy kernel, von Mises kernel, Directional data.

1. Introduction

Directional data are represented by an angle ($\theta \in [0, 2\pi), \theta = \theta + 2m\pi,\ m \in \mathbb{Z}$) or a unit vector ($\boldsymbol{x} = (\cos(\theta), \sin(\theta))^T$). Some examples of directional data are wind directions and electric power over a period of 24 hours and so on. For the preceding work of the kernel density estimation on the circle; See Hall et al. (1987), Taylor (2008), and Di Marzio et al. (2009, 2011).

Di Marzio et al. (2011) introduced the moments of sin-order p for the kernel on the circle. Its idea is somewhat similar to the moments of order p on the real line. The kernel of sin-order 2 includes the von Mises (VM) kernel, the wrapped Cauchy (WC) kernel and the wrapped normal kernel. Di Marzio et al. (2011) derived asymptotic mean integrated square error (AMISE) via the definition of the kernel functions with sin-order p. They show the convergent rate of the VM kernel is $O(n^{-4/5})$. This rate is equivalent to the convergent rate of the kernels of second order on the real line. However, the convergent rate of the AMISE can not be derived without specifying the kernel.

The VM kernel is called circular normal kernel. However, the VM kernel does not hold the reproductive property (Mardia and Jupp (1999, p44)). This is a big difference from the normal kernel on the line. The WC kernel holds the reproductive property and have more heavy tails than the VM kernel (Mardia and Jupp (1999, p39 and p52)). We explore properties of the WC kernel with respect to those advantages.

We show that the WC kernel has the optimal rate $O(n^{-2/3})$ of the AMISE and the asymptotic normality. Di Marzio et al. (2011) mentioned the another expression about

[*] Graguate School of Human and Socio-Environment Studies, Kanazawa University Kakumamachi Kanazawa 920-1192 Japan. tsuruta9764@gmail.com

[†] School of Economics, Kanazawa University Kakumamachi Kanazawa 920-1192 Japan. sagae.masahiko@gmail.com

the asymptotic normality of the circular kernel function. The WC kernel is inferior to the VM kernel with respect to the theoretical aspect of the MISE. However, our numerical experiments show that the WC kernel is superior to the VM kernel under the condition of the multimodal and/or the heavy tail.

2. Properties of AMISE of a circular kernel density estimation

Following Di Marzio et al. (2009, 2011), this section briefly refers definitions of the kernel function and properties of the AMISE.

2.1. Definition

Definiton 1 *(Kernel function)*

Let $K_\kappa(\theta)$ be a circular kernel function and κ be a concentration parameter (κ is a smoothing parameter corresponding to the inverse of bandwidth on the real line), where $K_\kappa(\theta) : [0, 2\pi) \to \mathbb{R}$ is such that

(a) It admits an uniformly convergent Fourier series,

$$K_\kappa(\theta) = (2\pi)^{-1}\{1 + \sum_{j=1}^{\infty} \gamma_j(\kappa)\cos(j\theta)\}, \quad \theta \in [0, 2\pi),$$

where $\gamma_j(\theta)$ is strictly monotonic function of κ,

(b) $\int_0^{2\pi} K_\kappa = 1$, and if $K_\kappa(\theta)$ takes negative values, there exists $0 < M < \infty$ such that, $\int_0^{2\pi} |K_\kappa| d\theta \le M$, for all $\kappa > 0$,

(c) $\lim_{\kappa\to\infty} \int_\delta^{2\pi-\delta} |K_\kappa(\theta)| d\theta = 0$, for all $0 < \delta < \pi$.

Definiton 2 *(Sin-order moment of the circular kernel)*

Let $\eta_j(K_\kappa) := \int_0^{2\pi} \sin^j(\theta) K_\kappa(\theta) d\theta$. K_κ of sin-order p is chosen so that,

$$\eta_0(K_\kappa) = 1, \qquad \eta_j(K_\kappa) = 0, \quad 0 < j < p, \qquad \text{and} \quad \eta_p(K_\kappa) \neq 0.$$

If $\eta_j(K_\kappa)$ is $j = 2s (s = 1, 2, \ldots)$, then $\eta_j(K_\kappa)$ is also represented by the sum of Fourier series $\gamma_j(\kappa)$:

$$\eta_j(K_\kappa) = \frac{1}{2^{2s-1}} \left[\binom{j-1}{j/2} + \sum_{s=1}^{j/2} (-1)^{j+s} \binom{j}{j/2+s} \gamma_{2s}(\kappa) \right]. \tag{1}$$

Definiton 3 *(Kernel density estimator)*

Let $\Theta_1, \ldots, \Theta_n$ be random sample from the unknown circular density $f(\theta)$. Given a circular kernel K_κ, the kernel estimator of f is defined as,

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

2.2. Theoretical properties

Theorem 1 *(AMISE)*
Under the following conditions:

(A) κ *increases as* $n \to \infty$. *for each* $j \in \mathbb{Z}^+$,$\lim_{n\to\infty} \gamma_j(\kappa) = 1$,

(B) $\lim_{n\to\infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\kappa) = 0$,

(C) f''*is continuous and square-integrable,*

the second sin-order kernel has,

$$AMISE[\hat{f}(\cdot;\kappa)] = \frac{\eta_2^2(K_\kappa)}{4} R\left(f''\right) + \frac{R\left(K_\kappa\right)}{n}, \tag{2}$$

where $\int_0^{2\pi} \{g(\theta)\}^2 d\theta = R(g)$, $\eta_2(K_\kappa) = \frac{1}{2}(1-\gamma_2(\kappa))$ *and* $R(K_\kappa) = (1+2\sum_{i=1}^{\infty} \gamma_j^2(\kappa))/(2\pi)$. *The right-hand side terms of* (2) *correspond to ISB and IV, respectively.*

The convergent rate of AMISE can not be derived without choosing specific kernel, since (2) depends on the kernel function K_κ. Di Marzio et al. (2011) derived AMISE of the VM kernel.

The VM kernel is defined as,

$$K_\kappa(\theta) := \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos\theta\},\ 0 < \kappa < \infty,$$

where $I_p(\kappa)$ denotes the modified Bessel function of the first kind with order p. The characteristic function of VM kernel is defined as,

$$\phi_p = \frac{I_p(\kappa)}{I_0(\kappa)} \quad p = 0, \pm 1, \ldots, \pm n, \cdots .$$

Proposition 1 *(AMISE of VM kernel)*
Noting $\gamma_j(\kappa) = I_j(\kappa)/I_0(\kappa)$, *the AMISE of the VM kernel is given as the following form :*

$$AMISE_{VM}[\hat{f}(\cdot;\kappa)] = \frac{1}{16}\left\{1 - \frac{I_2(\kappa)}{I_0(\kappa)}\right\}^2 R\left(f''\right) + \frac{1 + 2\sum_{j=1}^{\infty}\{I_j(\kappa)/I_0(\kappa)\}^2}{2n\pi}. \tag{3}$$

Under the condition that κ *is sufficiently large, the asymptotic form* (3) *reduces,*

$$AMISE_{VM}[\hat{f}(\cdot;\kappa)] = \frac{1}{4\kappa^2} R\left(f''\right) + \frac{\kappa^{1/2}}{2n\pi^{1/2}}. \tag{4}$$

We obtain the optimum κ^*:

$$\kappa^* = \left(2\pi^{1/2} R\left(f''\right) n\right)^{2/5}. \tag{5}$$

From (4) and (5), $\mathrm{AMISE}_{\mathrm{VM}} = O(n^{-4/5})$ is derived. Note that that $O(n^{-4/5})$ is equivalent to the convergent rate of the second-order kernels on the real line.

3. Properties of the Wrapped Cauchy kernel

The WC kernel is defined as,

$$K_\rho(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta)},\ 0<\rho\leq 1,$$

where ρ is concentration parameter. The characteristic function of the WC kernel is equal to,

$$\phi_p = \rho^{|p|}, \quad p = 0, \pm 1, \ldots, \pm n, \cdots .$$

Theorem 2 *(AMISE of WC kernel)*
The AMISE of the WC kernel is given as the follows,

$$AMISE_{WC}[\hat{f}(\cdot;\rho)] = \frac{\{1-\rho^2\}^2 R\left(f''\right)}{16} + \frac{1}{n\pi(1-\rho^2)}. \tag{6}$$

Let be $1-\rho^2 = h,\ 0 \leq h < 1$. *Then,* (6) *is expressed as,*

$$AMISE_{WC}[\hat{f}(\cdot;h)] = \frac{h^2 R\left(f''\right)}{16} + \frac{1}{n\pi h}. \tag{7}$$

See Appendix-A for the details.

In the similar way of Propositon 1. we obtain the optimum h^*:

$$h^* = \left(\frac{8}{\pi R\left(f''\right) n}\right)^{1/3}, \tag{8}$$

under the condition of $n > 8(\pi R\left(f''\right))^{-1}$. The optimal following convergent rate of $\mathrm{AMISE}_{\mathrm{WC}}$ is given by (7) and (8) as n is sufficiently large:

$$\mathrm{AMISE}_{\mathrm{WC}} = O(n^{-2/3}). \tag{9}$$

We should use ρ^* which minimize (6) as the practically optimal concentrating parameter, since (8) is larger than one if $n \leq 8(\pi R\left(f''\right))^{-1}$. ρ^* is given (7) and (8) as follows,

$$\rho^* = \arg\min_{0<\rho\leq 1}\left\{\mathrm{AMISE}_{\mathrm{WC}}[\hat{f}(\cdot;\rho)]\right\}. \tag{10}$$

We put $\hat{f}_\rho$ as $\hat{f}_h$ with $h = (1-\rho^2)$.

Theorem 3 *(Asymptotic Normality)*
Let be $\Theta_1, \Theta_2, \cdots \Theta_n \overset{i.i.d.}{\sim} f(\theta)$, $h = cn^{-\alpha}$ *and* $0 \leq h < 1$ *If* $\alpha > 1/3$, *then*

$$\sqrt{nh}[\hat{f}_h(\theta) - f(\theta)] \xrightarrow{d} N(0, f(\theta)/\pi), \quad n \to \infty. \tag{11}$$

See Appendix-B for the details.

The optimal convergent rate $O(n^{-2/3})$ of the AMISE of the WC kernel differs from the rate $O(n^{-4/5})$ of the VM kernel, although Both the WC kernel and the VM kernel are the second sin-order kernel. The Cauchy kernel on the real line has the

optimal convergent rate $O(n^{-2/3})$ by Davis (1975). The rate $O(n^{-2/3})$ of the WC kernel correspond to the rates of kernel family of order 0 such as the histogram and the Cauchy kernel.

We consider that the corespondence between the characteristic functions of the WC kernel and the Cauchy kernel causes the correspondence between the two rates. Mardia and Jupp (1999, p48 (3.5.59)) said the characteristic function of the WC Kernel corresponds that of Cauchy kernel $\phi(p) = e^{-a|p|}$. However, each of the inversion theorem is different. The inversion theorem of the WC kernel is Fourier series expansion, while that Cauchy kernel is Fourier transform.

4. Simulation

The optimal concentration parameter depends on $R(f'')$. The plug-in rule is the procedure to estimate the optimal concentration parameter with $\hat{R}(f'')$ as an estimator for $R(f'')$. The simplest rule among some plug-in rules is the procedure to assume that the true density f is the von Mises density f_{VM}. The simplest plug-in rule uses $\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa}))$ as the estimator for $R(f'')$, where κ is the concentration parameter of f_{VM} and $\hat{\kappa}$ is the maximum likelihood estimator of κ. $\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa}))$ corresponds to the following form:

$$\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa})) = \frac{\hat{\kappa}[3\hat{\kappa}I_2(2\hat{\kappa}) + 2I_1(2\hat{\kappa})]}{8\pi I_0^2(\hat{\kappa})}. \tag{12}$$

The normal procedure of plug-in rule is shown as,

i) Estimate $\hat{\kappa}$ from the sample and calculate $\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa}))$ from $\hat{\kappa}$,

ii) WC kernel: Estimate ρ^* which minimize (6) by substituting $\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa}))$ for $R(f'')$ in (6),

iii) VM kernel : Estimate κ^* by substituting $\hat{R}(f''_{\mathrm{VM}}(\cdot;\hat{\kappa}))$ for $R(f'')$ in (5).

The two simulations are given as,
Experiment-A (mixture of the von Mises density)

1. The WC kernel:
   (a) Let a true density be a mixture of the von Mises density:

$$f(\theta)_{\mathrm{MVM}} = \frac{1}{2}f_{\mathrm{VM1}}(\theta;\mu_1,\kappa_1) + \frac{1}{2}f_{\mathrm{VM2}}(\theta;\mu_2,\kappa_2), \tag{13}$$

   where let be $\mu_1 = \pi/2$, $\mu_2 = 3\pi/2$ and $\kappa_1 = \kappa_2 = \kappa$. Generate the random sample of the size n distributed as (13),
   (b) Estimate ρ^* by the plug-in rule,
   (c) Let $\mathrm{ISE} = \int_0^{2\pi}\{\hat{f}(\theta;\rho) - f(\theta)\}^2 d\theta$, Let $\overline{\mathrm{ISE}}(\hat{f}(\cdot;\rho))$ be the numerical integration of ISE. Calculate $\overline{\mathrm{ISE}}(\hat{f}(\cdot;\rho))$,
   (d) Repeat (a)-(c) 1000 times and compute $\overline{\mathrm{MISE}}(\hat{f}(\cdot;\rho)) = \sum_{i=1}^{1000}\overline{\mathrm{ISE}}_i(\hat{f}(\cdot;\rho))/1000$.
2. The VM kernel :
   With the same procedures as (a) - (d) of the WC kernel, compute $\overline{\mathrm{MISE}}(\hat{f}(\cdot;\kappa^*))$,

3. Calculate $\overline{\mathrm{MISE}}(\hat{f}(\cdot;\kappa^*)) - \overline{\mathrm{MISE}}(\hat{f}(\cdot;\rho^*))$, where it round off decimal point forth place.

Experiment-B (wrapped Cauchy density)

1. Let a true density be a wrapped Cauchy density given by,

$$f_{wc}(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)},$$

where directional mean $\mu = \pi$.

2. Calculate $\overline{\mathrm{MISE}}(\hat{f}(\cdot;\kappa^*)) - \overline{\mathrm{MISE}}(\hat{f}(\cdot;\rho^*))$ in the same procedures as Experiment-A.

Table 1: $\overline{\mathrm{MISE}}(\hat{f}_{\kappa^*}) - \overline{\mathrm{MISE}}(\hat{f}_{\rho^*})$. n is sample size, κ is represented as concentrate parameter of true density f, which is a mixture of von Mises density.

	$n=10$	50	100	200	500	1000
$\kappa=0.3$	0.011	0.002	0.001	0	0	0
0.5	0.012	0.002	0.001	0	0	0
0.7	0.011	0.002	0.001	0	0	0
1	0.012	0.001	0	0	0	0
2	0.007	-0.004	-0.005	-0.005	-0.005	-0.005
5	-0.017	-0.023	-0.025	-0.024	-0.024	-0.023
10	-0.032	-0.036	-0.037	-0.037	-0.037	-0.037
15	-0.037	-0.044	-0.043	-0.042	-0.042	-0.042
20	-0.039	-0.045	-0.045	-0.045	-0.048	-0.046

Table 2: $\overline{\mathrm{MISE}}(\hat{f}_{\kappa^*}) - \overline{\mathrm{MISE}}(\hat{f}_{\rho^*})$. n is sample size, ρ is represented as concentrate parameter of true density f, which is a wrapped Cauchy density.

	n =50	100	200	500	1000
ρ =0.5	-0.015	-0.009	-0.006	-0.003	-0.002
0.7	-0.011	-0.006	-0.003	-0.001	-0.001
0.8	0.007	0.012	0.013	0.011	0.009
0.9	0.133	0.141	0.134	0.112	0.093

The result of Table 1 indicates that the WC kernel is superior to the VM kernel under that n is not sufficiently large ($n \leq 100$) and f has the mixture of the von Mises (the multimodal) and $\kappa \leq 2$. This difference between the WC kernel and the VM kernel tends to become smaller as n become larger.

The result of Table 2 shows that WC kernel has better than the VM kernel if $n \leq 1000$ and f has the wrapped Cauchy (the heavy tail) and $\rho \geq 0.8$. Particularly, the WC kernel has very better behavior about the top of f than the VM kernel even if n is large.

5. Conclusion

This paper shows that the WC kernel has the convergent rate $O(n^{-2/3})$ of the AMISE and the asymptotic normality. However, the convergent rate of the VM kernel is $O(n^{-4/5})$. In other words, the rate of the WC kernel and that of the VM kernel do not equal in spite of the same second sin-order kernel.

The convergent rate of the AMISE of the WC kernel corresponds to the rate of Cauchy kernel, since the characteristic function of the WC kernel corresponds to the one of Cauchy kernel.

The WC kernel is not better than the VM kernel with respect to the AMISE convergent rate. However, the results of our simulations show that the WC kernel is superior to the VM kernel if a true density f has the multimodal and/or the heavy tail.

6. Acknowledgements

I would like to thank the editor and the referee for the helpful comments.This work was supported by JSPS KAKENHI Grant Number 16K00043.

7. Appendix

7.1. Appendix-A

PROOF OF THEOREM 2. The WC kernel has $\gamma_j(\rho) = \rho^j$ and $\eta_2(K_\rho) = (1-\rho^2)/2$. Since for small values of $u, \sin(u) \simeq u$, we use the expansion $f(\theta+u) \simeq f(\theta) + f'(\theta)\sin(u) + f''(\theta)\sin^2(u)/2 + O(\sin^3(u))$.

$$\begin{aligned}
\mathrm{E}_f[K_\rho(\theta - Y)] &= \int K_\rho(\theta - y) f(y) dy \\
&= \int K_\rho(u) f(\theta + u) du \\
&= \int K_\rho(u)[f(\theta) + f'(\theta)\sin(u) + f''(\theta)\sin^2(u)/2 + O(\sin^3(u))] \\
&= f(\theta) + \frac{1}{2}\eta_2(K_\rho) f''(\theta) + o(1) \\
&= f(\theta) + \frac{1}{4}(1-\rho^2) f''(\theta) + o(1). \qquad (14)
\end{aligned}$$

$$\begin{aligned}
R(K_\rho) &= \frac{1 + 2\sum_{j=1}^{\infty}(\rho^j)^2}{2\pi} \\
&= \frac{1 + \frac{2\rho^2}{1-\rho^2}}{2\pi} \\
&= \frac{1+\rho^2}{2\pi(1-\rho^2)} \\
&= \frac{1}{\pi(1-\rho^2)} - \frac{1}{2\pi}. \qquad (15)
\end{aligned}$$

It follows from (15) that

$$\begin{aligned}
n^{-1}\mathrm{Var}_f[K_\rho(\theta - Y)] &= n^{-1}\left\{E_f[K_\rho^2(\theta - Y)] - \mathrm{E}_f[K_\rho(\theta - Y)]^2\right\} \\
&= n^{-1}\left[\frac{1}{\pi(1-\rho^2)} - \frac{1}{2\pi}\right]\{f(\theta) + o(1)\} - n^{-1}[f(\theta) + o(1)] \\
&= \frac{f(\theta)}{\pi n(1-\rho^2)} + o(n^{-1}(1-\rho^2)^{-1}). \qquad (16)
\end{aligned}$$

7.2. Appendix-B

PROOF OF THEOREM 3. Put $h = (1-\rho^2)$, Write (14) and (16) as,

$$\mathrm{E}_f[K_h(\theta - \Theta_1)] \simeq f(\theta) + \frac{1}{4}hf''(\theta) + o(1), \qquad (17)$$

$$\mathrm{Var}_f[K_h(\theta - \Theta_1)] = \frac{f(\theta)}{\pi h} + o(h^{-1}). \qquad (18)$$

$$\sqrt{nh}[\hat{f}_h(\theta) - f(\theta)] = \sqrt{nh}\{\hat{f}_h(\theta) - \mathrm{E}[\hat{f}_h(\theta)]\} + \sqrt{nh}\mathrm{bias}[\hat{f}_h(\theta)]. \qquad (19)$$

The first term of (19) is equal to,

$$\sqrt{nh}\{\hat{f}_h(\theta) - \mathrm{E}_f[\hat{f}_h(\theta)]\} = \sqrt{n}\left\{n^{-1}\sum_{i=1}^{n} h^{1/2}K_h(\theta-\Theta_i) - \mathrm{E}_f[h^{1/2}K_h(\theta-\Theta_1)]\right\}. \tag{20}$$

$$\begin{aligned}\mathrm{E}_f[h^{1/2}K_h(\theta-\Theta_1)] &= h^{1/2}\mathrm{E}_f[K_h(\theta-\Theta_1)] \\ &\simeq h^{1/2}\left[f(\theta) + \frac{f^{(\prime\prime)}(\theta)}{4}h\right].\end{aligned} \tag{21}$$

It is shown that $0 \leq |\mathrm{E}[h^{1/2}K_h(\theta-\Theta_1)]| < \infty$ from (21).

From (18), We obtain the following form:

$$\begin{aligned}\mathrm{Var}_f[h^{1/2}K_h(\theta-\Theta_1)] &= h\mathrm{Var}[K_h(\theta-\Theta_1)] \\ &= h\left[\frac{f(\theta)}{\pi h} + o(h^{-1})\right] \\ &= \frac{f(\theta)}{\pi} + o(1) \\ &\simeq \frac{f(\theta)}{\pi}.\end{aligned} \tag{22}$$

It is shown that $\mathrm{Var}_f[h^{1/2}K_h(\theta-\Theta_1)] < \infty$ from (22).

Since (20) satisfies the condition of Lindeberg (Feller (1968, p.244)) from (21) and (22), it is given as the follows,

$$\sqrt{nh}\{\hat{f}_h(\theta) - \mathrm{E}_f[\hat{f}_h(\theta)]\} \xrightarrow{d} \mathrm{N}(0, f(\theta)/\pi), \quad n \to \infty. \tag{23}$$

The order of the second term of (19) is equal to,

$$\begin{aligned}\sqrt{nh}\mathrm{bias}[\hat{f}_h] &= \sqrt{nh}O(h) \\ &= O(\sqrt{nh^3}).\end{aligned} \tag{24}$$

with $h = cn^{-\alpha}$, we obtain the follows,

$$\begin{aligned}\sqrt{nh^3} &\sim n^{1/2}n^{-3\alpha/2} \\ &= n^{(1-3\alpha)/2}.\end{aligned}$$

When $\alpha > 1/3$ is chosen, (24) is given as the following form:

$$\begin{aligned}\sqrt{nh}\mathrm{bias}[\hat{f}_h(\theta)] &= O(\sqrt{nh^3}) \\ &= o(1).\end{aligned} \tag{25}$$

For $\alpha > 1/3$ and as $n \to \infty$, Theorem 3 completes the proof from (23) and (25).

References

Davis, K. B. (1975). Mean Square Error Properties of Density Estimates, *The Annals of Statistics* **3**, 1025-1030.

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

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

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

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

Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley.

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

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley.

Received June 11, 2016
Revised January 16, 2017