

# PROPERTIES FOR CIRCULAR NONPARAMETRIC REGRESSIONS BY VON MIESE AND WRAPPED CAUCHY KERNELS

Tsuruta, Yasuhito  
Wakayama prefectural office

Sagae, Masahiko  
School of Economics, Kanazawa University

<https://doi.org/10.5109/2232334>

---

出版情報 : Bulletin of informatics and cybernetics. 50, pp.1-13, 2018-12. Research Association  
of Statistical Sciences

バージョン :

権利関係 :



PROPERTIES FOR CIRCULAR NONPARAMETRIC REGRESSIONS  
BY VON MIESE AND WRAPPED CAUCHY KERNELS

by

Yasuhito TSURUTA    and    Masahiko SAGAE

---

*Reprinted from the Bulletin of Informatics and Cybernetics  
Research Association of Statistical Sciences, Vol.50*

---

FUKUOKA, JAPAN  
2018

# PROPERTIES FOR CIRCULAR NONPARAMETRIC REGRESSIONS BY VON MIESE AND WRAPPED CAUCHY KERNELS

By

Yasuhito TSURUTA\* and Masahiko SAGAE†

## Abstract

We discuss the asymptotic properties with respect to nonparametric regression for circular data. We reveal theoretical properties for circular nonparametric regression by applying von Mises (VM) and wrapped Cauchy (WC) kernels. We derive the asymptotic normalities and the convergence rate of the weighted conditional mean integrated squared errors regarding VM and WC kernels. The numerical experiment shows that WC kernel outperforms VM kernel in the small samples, and the theoretical properties are supported in the large samples.

*Key Words and Phrases:* Nonparametric regression, Circular data, Von Mises distribution, Wrapped Cauchy distribution.

## 1. Introduction

We aim to model the relation between a linear response variable  $Y_i$  for  $y_i \in \mathbb{R}$  and an circular explanatory variable  $\Theta_i$  for  $\theta_i \in [-\pi, \pi)$ . Let the data set  $\{(Y_1, \Theta_1), \dots, (Y_n, \Theta_n)\}$  be i.i.d.. Then, we consider that

$$Y_i = m(\Theta_i) + v^{1/2}(\Theta_i)\varepsilon_i,$$

where  $v(\theta) = \text{Var}_Y[Y|\Theta = \theta]$  is the conditional variance,  $\varepsilon_i$  is a random variable on the real line with zero mean and unit variance, and a regression function  $m(\theta) := E_Y[Y|\Theta = \theta]$  is periodic such as  $m(\theta) = m(\theta + 2\pi)$ .

We consider a regression being able to estimate  $m(\theta)$  under less rigid assumptions. One of the estimator is a nonparametric regression. In nonparametric regressions for circular data analysis, a sine local linear regression (S-LLR)  $\hat{m}(\theta; \kappa)$  is proposed by Di Marzio et al. (2009). S-LLR  $\hat{m}(\theta; \kappa)$  is defined as  $\hat{\beta}_0$  in  $(\hat{\beta}_0, \hat{\beta}_1)^T$  such that minimizes

$$\sum_{i=1}^n \{Y_i - \beta_0 - \beta_1 \sin(\Theta_i - \theta)\}^2 K_\kappa(\Theta_i - \theta). \quad (1)$$

where  $K_\kappa(\theta_i - \theta)$  is a symmetric kernel function, and  $\kappa$  is a concentration parameter that is a smoothing parameter corresponding to the inverse of the squared bandwidth:  $\kappa = h^{-2}$ .

---

\* Wakayama prefectural office, 1-1 Komatsubaradori, Wakayama-shi, Wakayama-ken, 640-8585, tsuruta9764@gmail.com

† School of Economics, Kanazawa University, Kakuma-machi Kanazawa-shi, Ishikawa-ken, 920-1192, Japan. sagae.masahiko@gmail.com

Di Marzio et al. (2009) derived the conditional mean squared error (MSE) of S-LLR by employing von Mises (VM) kernel, and calculated the optimal parameter of this and the convergence rate of this MSE. However, few study explored the theoretical properties for S-LLR employing another kernel functions such as wrapped Cauchy (WC) kernel, and no study shown the global properties of S-LLR such as the conditional weighted mean integrated squared error (MISE) as far as we know. Accordingly, we elucidate the MISEs of VM and WC kernels.

In section 2 we show the definitions of S-LLR and a class of kernels, and explain the MSE of S-LLR by Di Marzio et al. (2009). In addition we prove the asymptotic normality for S-LLR. This result can provide the confidence interval for S-LLR.

In section 3 we derive the MISE and the asymptotic normality of VM kernel. We show that the convergence rate of this MISE is  $O_p(n^{-4/5})$ .

In section 4 we provide the MISE and the asymptotic normality of WC kernel. We show that the rate of this MISE is  $O_p(n^{-2/3})$ . In the study of kernel density estimations for circular data, Tsuruta and Sagae (2017) derived that the rate of the MISE of VM kernel is  $O(n^{-4/5})$ , and that of WC kernel is  $O(n^{-2/3})$ .

In section 5 we conduct the numerical experiment to compare the both performances under finite samples. This experiment demonstrates that WC kernel exhibit better properties than VM kernel when the sample is small, and VM kernel well performs than WC kernel when the sample is large enough.

## 2. Sine local linear regression (S-LLR)

S-LLR  $\hat{m}(\theta; \kappa)$  provided by minimizing (1) is given by

$$\hat{m}(\theta; \kappa) := \mathbf{e}_1^T (\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{Y}, \quad (2)$$

where  $\mathbf{e}_1$  is the  $2 \times 1$  vector having 1 in the first and zero elsewhere,  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  is the vector of the responses,

$$\mathbf{S}_\theta := \begin{bmatrix} 1 & \sin(\Theta_1 - \theta) \\ \vdots & \vdots \\ 1 & \sin(\Theta_n - \theta) \end{bmatrix}$$

is an  $n \times 2$  design matrix, and  $\mathbf{W}_\theta := \text{diag}\{K_\kappa(\Theta_1 - \theta), \dots, K_\kappa(\Theta_n - \theta)\}$  is an  $n \times n$  diagonal matrix. Additionally, we assume that the inverse  $(\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1}$  exists.

We employ a class of kernels  $K_\kappa(\theta)$  satisfying the following definition proposed by Di Marzio et al. (2009) and Di Marzio et al. (2011).

**Definition 1** *The kernel  $K_\kappa(\theta)$  is a non-negative function satisfying the two following conditions:*

(a) *It admits a convergent Fourier series representation:*

$$K_\kappa(\theta) = 1/(2\pi) \left\{ 1 + 2 \sum_{j=1}^{\infty} \gamma_j(\kappa) \cos(j\theta) \right\},$$

where  $\gamma_j(\kappa) := E_K[\cos(j\theta)]$  and  $\gamma_j(\kappa)$  are monotonic functions of  $\kappa$ .

(b) For all  $0 < \delta < \pi$ ,  $\lim_{\kappa \rightarrow \infty} \int_{\delta \leq |\theta| \leq \pi} |K_\kappa(\theta)| d\theta = 0$ .

We now define a  $j$ th sine-type moment as

$$\eta_j(K_\kappa) := \int_{-\pi}^{\pi} \sin(\theta)^j K_\kappa(\theta) d\theta$$

The  $j$ th sine-type moment  $\eta_j(K_\kappa)$  plays a similar role as a  $j$ th moment of a symmetric kernel on the real line. Especially, the second sine-type moment is given by

$$\eta_2(K_\kappa) = (1 - \gamma_2(\kappa))/2. \quad (3)$$

Put  $\Theta_n := \{\Theta_1, \dots, \Theta_n\}$  and  $R(g) := \int_{-\pi}^{\pi} g(\theta)^2 d\theta$ . Then, let the conditional bias be  $\text{Bias}_Y[\hat{m}(\theta; \kappa) | \Theta_n] =: E_Y[\hat{m}(\theta; \kappa) | \Theta_n] - m(\theta)$  and the conditional variance of S-LLR be  $\text{Var}_Y[\hat{m}(\theta; \kappa) | \Theta_n]$ . Di Marzio et al. (2009) derived the following theorem regarding the bias and the variance.

**Theorem 1** *Assume that the following four conditions hold:*

- i)  $\lim_{n \rightarrow \infty} n^{-1} R(K_\kappa) = 0$ .
- ii)  $\lim_{n \rightarrow \infty} \gamma_j(\kappa) = 1$ .
- iii) *The marginal density  $f(\theta)$  is continuously differentiable, where  $f(\theta) > 0$  for any  $\theta$ .*
- iv) *The second derivative  $m''(\theta)$  and the conditional variance  $v(\theta)$  are continuous, respectively.*

Then, the bias is approximately given by

$$\text{Bias}_Y[\hat{m}(\theta; \kappa) | \Theta_n] \simeq \eta_2(K_\kappa) \frac{m''(\theta)}{2!}, \quad (4)$$

and the variance is approximately given by

$$\text{Var}_Y[\hat{m}(\theta; \kappa) | \Theta_n] \simeq R(K_\kappa) \frac{v(\theta)}{nf(\theta)}. \quad (5)$$

We derive the following asymptotic normality of S-LLR from combining Theorem 1 and Lindeberg's central limit theorem (CLT).

**Theorem 2** *Assume that the all conditions of Theorem 1 hold. Then, it follows that*

$$\sqrt{n/R(K_\kappa)} [\hat{m}(\theta; \kappa) - E_Y[\hat{m}(\theta; \kappa) | \Theta_n]] \xrightarrow{d} N(0, v(\theta)/f(\theta)) \quad n \rightarrow \infty.$$

The proof is presented in Appendix A.

We find out that the bias depends on  $\eta_2(K_\kappa)$ , and the variance depends on  $R(K_\kappa)$  in Theorem 1. For providing the convergence rate of the MSE in Theorem 1, it is needed to divide this two terms into  $\kappa$  and any constant part  $C(K)$ , but it is difficult to obtain general kernel's conditions enabling this dividing. Therefore, we choose the two well-used VM and WC kernels in circular data, and derive the asymptotic properties for S-LLR applying the two kernels.

### 3. Theoretical properties for von Mises kernel

VM kernel  $K_\kappa(\theta)$  is defined as

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

where  $I_p(\kappa)$  denotes the modified Bessel function of the first kind and order  $p$ . The coefficients of VM kernel are given by

$$\gamma_j(\kappa) = I_j(\kappa)/I_0(\kappa). \quad (6)$$

VM kernel (density) is called as the circular normal density for having the properties being similar to the normal density. For example, VM kernel has good properties such that this belongs to exponential family, and the maximum likelihood estimators (MLEs) of this have the explicit solutions. However, VM kernel does not satisfy the reproductive property.

The second sine-type moment  $\eta_2(K_\kappa)$  and the term  $R(K_\kappa)$  of VM kernel are presented in the following Lemma.

**Lemma 1** *From combining (3) and (3.5.37) in Mardia and Jupp (1999), the second sine-type moment for VM kernel is given by*

$$\eta_2(K_\kappa) = \frac{I_1(\kappa)}{\kappa I_0(\kappa)}. \quad (7)$$

*If  $\kappa$  is large enough, then from combining (7) and (3.5.34) in Mardia and Jupp (1999), the second sine-type moment  $\eta_2(K_\kappa)$  is equal to*

$$\eta_2(K_\kappa) = \frac{1}{\kappa} \{1 + o_p(1)\}.$$

*From (3.5.27) in Mardia and Jupp (1999), the term  $R(K_\kappa)$  is given by*

$$R(K_\kappa) = \frac{I_0(2\kappa)}{2\pi I_0(\kappa)^2}. \quad (8)$$

*If  $\kappa$  is large enough, then from combining (8) and (3.5.33) in Mardia and Jupp (1999), the term  $R(K_\kappa)$  can approximate to*

$$R(K_\kappa) \simeq \kappa^{1/2}/(2\pi^{1/2}).$$

We define MISE as  $\text{MISE}_Y[\hat{m}(\theta; \kappa) | \Theta_n] := \text{E}_Y[\int_{-\pi}^{\pi} \{\hat{m}(\theta; \kappa) - m(\theta)\}^2 f(\theta) d\theta | \Theta_n]$ . Then, we obtain the following theorem with respect to VM kernel from combining Theorem 1 and Lemma 1.

**Theorem 3** *Assume that as  $n \rightarrow \infty$ ,  $\kappa \rightarrow \infty$ , and  $n^{-1}\kappa^{1/2} \rightarrow 0$ . Then, the bias is approximately given by*

$$\text{Bias}_Y[\hat{m}(\theta; \kappa) | \Theta_n] \simeq \frac{1}{2\kappa} m''(\theta), \quad (9)$$

and the variance is approximately given by

$$\text{Var}_Y[\hat{m}(\theta; \kappa) | \Theta_n] \simeq \frac{\kappa^{1/2} v(\theta)}{2\pi^{1/2} n f(\theta)}. \quad (10)$$

From combining (9) and (10), we obtain the following asymptotic MISE that is

$$\text{AMISE}_Y[\hat{m}(\theta; \kappa) | \Theta_n] = \frac{1}{4\kappa^2} \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta + \frac{\kappa^{1/2} \int_{-\pi}^{\pi} v(\theta) d\theta}{2\pi^{1/2} n}. \quad (11)$$

The minimizer  $\kappa_*$  of (11) is given by

$$\kappa_* = \left[ \frac{2\pi^{1/2} \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta}{\int_{-\pi}^{\pi} v(\theta) d\theta} \right]^{2/5} n^{2/5}. \quad (12)$$

Therefore, the optimal  $\text{AMISE}_Y[\hat{m}(\theta; \kappa_*) | \Theta_n]$  is  $O_p(n^{-4/5})$ .

We obtain the following asymptotic normal distribution of VM kernel from Theorems 2 and 3, and Lemma 1.

**Theorem 4** Put  $\kappa = cn^\alpha$ , where  $c$  and  $\alpha$  are any constants. Then, if  $\alpha > 2/5$  and  $n \rightarrow \infty$ , then it holds that

$$n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - m(\theta)] \xrightarrow{d} N(0, v(\theta) / \{2\pi^{1/2} f(\theta)\}),$$

The proof is presented in Appendix B.

#### 4. Theoretical properties wrapped Cauchy kernel

WC kernel is defined as

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

where  $\rho$  is the concentration parameter. The coefficients of WC kernel are given by

$$\gamma_j(\rho) = \rho^j.$$

The coefficients  $\gamma_j(\rho)$  are very simpler forms than that of VM kernel. Note that WC kernel satisfies the reproductive property. This two points are advantages for WC kernel. However, the MLEs of WC kernel generally does not have the explicit solutions.

The second sine-type moment  $\eta_2(K_\rho)$  and the term  $R(K_\rho)$  of WC kernel are presented in the following Lemma.

**Lemma 2** The second sine-type moment of WC kernel is given by

$$\eta_2(K_\rho) = (1 - \rho^2)/2.$$

Using Parseval's formula:  $R(K_\rho) = (2\pi)^{-1} \{1 + 2 \sum_{j=1}^{\infty} \gamma_j(\rho)^2\}$ , the term  $R(K_\rho)$  is approximately equal to

$$\begin{aligned} R(K_\rho) &= \frac{1}{\pi(1 - \rho^2)} - \frac{1}{2\pi} \\ &= \frac{1}{\pi(1 - \rho^2)} \{1 + o_p(1)\}. \end{aligned}$$

We now put  $h = 1 - \rho^2$   $0 < h < 1$ . Then, we derive the bias, the variance and the MISE of WC kernel from combining Theorem 1 and Lemma 2.

**Theorem 5** *Assume that as  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $nh \rightarrow \infty$ . Then, the bias is approximately given by*

$$\text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n] \simeq h \frac{m''(\theta)}{4}, \quad (13)$$

and the variance is approximately given by

$$\text{Var}_Y[\hat{m}(\theta; h)|\Theta_n] \simeq (nh)^{-1} \frac{v(\theta)}{\pi f(\theta)}. \quad (14)$$

From combining (13) and (14), we obtain the asymptotic MISE that is

$$\text{AMISE}_Y[\hat{m}(\theta; h)|\Theta_n] = \frac{h^2 \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta}{16} + \frac{\int_{-\pi}^{\pi} v(\theta) d\theta}{\pi nh}. \quad (15)$$

The minimizer  $h_*$  of (15) is given by

$$h_* = \left\{ \frac{8 \int_{-\pi}^{\pi} v(\theta) d\theta}{\pi \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta} \right\}^{1/3} n^{-1/3}.$$

Hence, the optimal  $\text{AMISE}_Y[\hat{m}(\theta; h_*)|\Theta_n]$  is  $O_p(n^{-2/3})$ .

Comparing Theorems 3 and 5, these results indicate that S-LLR's convergence rates for AMISE are different speeds depending on an employed kernel. This is different from the property that the standard LLR on the real line always has the same rate of  $O_p(n^{-4/5})$  under non-negative kernels.

We obtain the following asymptotic distribution for WC kernel from combining Theorems 2 and 5, and Lemma 2.

**Theorem 6** *Put  $h = cn^\alpha$ , where  $c$  and  $\alpha$  are any constants. Then, if  $\alpha < -1/3$  and  $n \rightarrow \infty$ , then it holds that*

$$(nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)] \xrightarrow{d} N(0, v(\theta)/\{\pi f(\theta)\}),$$

## 5. Numerical experiment

We discussed the theoretical aspects for VM and WC kernels in the above sections. From practical view point we want to investigate the performances in small samples for the both kernel through a numerical experiment.

We consider

$$Y_i = m(\Theta_i) + v^{1/2}(\Theta_i)\varepsilon_i \quad \varepsilon_i \stackrel{i.i.d.}{\sim} N(0, 1), \quad \text{and} \quad v(\Theta_i) = t^2,$$

where we employ the two following regression functions:

- Function 1  $m_1(\theta) := 2 + 3 \cos(\theta) + 2 \sin(3\theta)$ .
- Function 2  $m_2(\theta) := 2 + 0.7 \cos(\theta) + 0.7^2 \cos(2\theta) + 0.7^3 \cos(3\theta)$ .

Figure 1 shows that Function 1 is asymmetric and has some waves, and Function 2 is symmetric and has a sharp mountain. Additionally, we employ a circular uniform distribution and VM distribution for the distribution of  $\Theta_i$ . In other words, we now consider four models in Table 1.

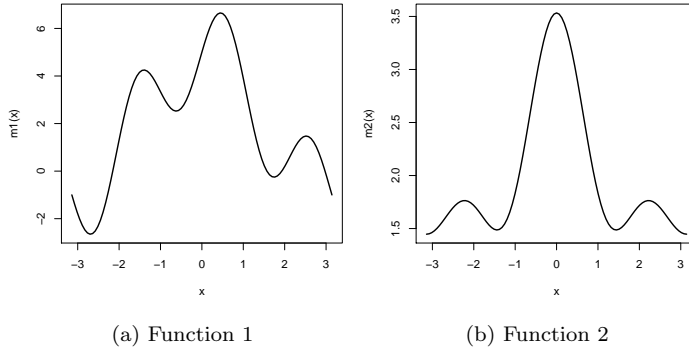


Figure 1: The true regression functions 1 and 2.

Table 1: Models 1–4 of the numerical experiment.

Model	Function	Distribution of $\Theta_i$
1	1	Circular uniform
2	1	Von Mises ( $\mu = 0, \kappa = 0.3$ )
3	2	Circular uniform
4	2	Von Mises ( $\mu = 0, \kappa = 0.3$ )

We apply the least squares cross validation (LSCV) to estimate the optimal concentration parameter. LSCV estimator employing VM kernel is given by

$$\hat{\kappa} := \arg \min_{\kappa > 0} \left[ \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}_{-i}(\Theta_i; \kappa)\}^2 \right], \quad (16)$$

where  $\hat{m}_{-i}(\theta; \kappa)$  is S-LLR given by removing  $i$ -th observation  $(Y_i, \Theta_i)$ . In additions, We provide LSCV estimator employing WC kernel  $\hat{\rho}$  given by replacing  $\kappa$  by  $\rho$  in (16). We conduct the numerical experiment according to the following procedure:

Execute the following six steps for each model:

1. Generate a random sample  $\{\Theta_1, \dots, \Theta_n\}$  following the distribution in Table 1.

2. Generate a random sample  $\{\varepsilon_1, \dots, \varepsilon_n\}$  following in the normal distribution  $N(0, t^2)$ .
3. Generate a random sample  $\{Y_1, \dots, Y_n\}$  from (a)–(b).
4. Give VM kernel  $\hat{m}(\theta; \hat{\kappa}_{CV})$  and WC kernel  $\hat{m}(\theta; \hat{\rho}_{CV})$  with each LSCV estimator.
5. Calculate the numerical integral  $ISE = \int_{-\pi}^{\pi} \{\hat{m}(\theta; \cdot) - m(\theta)\} f(\theta) d\theta$  for the above two estimators, where  $f(\theta)$  is the density of the distribution in (a).
6. Repeat 1000 times from (a)–(e), and calculate the means and standard divisions of ISE of the two estimators in (d).

Tables 2–5 show that WC kernel outperforms VM kernel when  $n$  is small ( $n \leq 20$  or  $n \leq 30$ ) for all models. Additionally, WC kernel is more stable when  $n \leq 50$ . In models 3–4, WC kernel always shows the better performances when  $n \leq 50$ . Even when  $n = 100$ , it outperforms VM kernel in  $t = 2$  in model 3. The result indicates that WC kernel has a advantage in small samples if a regression function has a sharp mountain. VM kernel better performs when  $n = 100$  for all models except in the case where  $t = 2$  in models 3–4. We recommend to apply WC kernel if  $n$  is small. However, we should employ VM kernel if  $n$  is large.

Table 2: Result of model 1. VM and WC represents VM and WC kernels employing in S-LLR with each LSCV estimator. The values are the means and standard divisions of the numerical ISE of each S-LLR, based on 1000 simulated samples of size  $n = 10, 20, 30, 40, 50$ , and 100 and standard divisions of the normal error  $t = 1$  (above the double rules) and  $t = 2$  (below the double rules).

$n$	VM		WC	
	MISE	sd(ISE)	MISE	sd(ISE)
10	16.32	305.69	2.77	1.62
20	3.24	44.22	1.37	0.77
30	0.66	0.73	0.88	0.43
40	0.44	0.30	0.68	0.31
50	0.34	0.18	0.55	0.23
100	0.16	0.06	0.30	0.09
10	11.86	103.37	4.34	2.26
20	3.38	15.84	2.48	1.19
30	1.62	0.91	1.76	0.77
40	1.23	0.73	1.41	0.61
50	0.99	0.51	1.20	0.51
100	0.50	0.22	0.70	0.24

Table 3: Result of model 2. VM and WC represents VM and WC kernels employing in S-LLR with each LSCV estimator. The values are the means and standard divisions of the numerical ISE of each S-LLR, based on 1000 simulated samples of size  $n = 10, 20, 30, 40, 50$ , and 100 and standard divisions of the normal error  $t = 1$  (above the double rules) and  $t = 2$  (below the double rules).

$n$	VM		WC	
	MISE	sd(ISE)	MISE	sd(ISE)
10	18.98	258.3	2.82	1.59
20	1.44	2.24	1.34	0.69
30	0.78	1.65	0.91	0.46
40	0.66	5.19	0.68	0.29
50	0.37	0.50	0.55	0.21
100	0.17	0.07	0.31	0.10
10	52.09	1032.93	4.38	2.11
20	2.88	4.84	2.45	1.08
30	1.88	4.84	1.77	0.77
40	1.28	0.94	1.40	0.56
50	1.02	1.29	1.18	0.47
100	0.51	0.23	0.70	0.25

Table 4: Result of model 3. VM and WC represents VM and WC kernels employing in S-LLR with each LSCV estimator. The values are the means and standard divisions of the numerical ISE of each S-LLR, based on 1000 simulated samples of size  $n = 10, 20, 30, 40, 50$ , and 100 and standard divisions of the normal error  $t = 1$  (above the double rules) and  $t = 2$  (below the double rules).

$n$	VM		WC	
	MISE	sd(ISE)	MISE	sd(ISE)
10	1.42	12.99	0.52	0.30
20	0.45	1.57	0.34	0.17
30	0.27	0.17	0.26	0.14
40	0.21	0.15	0.21	0.11
50	0.17	0.11	0.18	0.10
100	0.09	0.05	0.10	0.05
10	2.34	6.39	1.25	1.11
20	1.13	4.57	0.80	0.56
30	0.66	0.51	0.61	0.44
40	0.56	0.56	0.51	0.31
50	0.47	0.34	0.46	0.32
100	0.28	0.17	0.27	0.15

Table 5: Result of model 4. VM and WC represents VM and WC kernels employing in S-LLR with each LSCV estimator. The values are the means and standard divisions of the numerical ISE of each S-LLR, based on 1000 simulated samples of size  $n = 10, 20, 30, 40, 50$ , and 100 and standard divisions of the normal error  $t = 1$  (above the double rules) and  $t = 2$  (below the double rules).

$n$	VM		WC	
	MISE	sd(ISE)	MISE	sd(ISE)
10	2.18	27.7	0.52	0.31
20	0.40	0.38	0.34	0.18
30	0.31	1.15	0.26	0.14
40	0.33	3.59	0.20	0.11
50	0.17	0.11	0.18	0.10
100	0.09	0.05	0.10	0.05
10	6.85	94.68	1.23	1.08
20	0.95	1.15	0.81	0.58
30	0.87	4.60	0.65	0.49
40	0.57	0.44	0.52	0.34
50	0.49	0.36	0.47	0.32
100	0.28	0.19	0.28	0.18

## 6. Acknowledgements

We would like to thank the editor and reviewer for some helpful comments. This work was supported by JSPS KAKENHI Grant Numbers JP16K00043, JP24500339, and JP16H02790.

## References

- 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. (1966). *An Introduction to Probability Theory and Its Applications. II* (Second ed.). John Wiley & Sons, Inc.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. John Wiley & Sons, Inc.
- Tsuruta, Y., and Sagae, M. (2017). Asymptotic Property of Wrapped Cauchy Kernel Density Estimation on the Circle. *Bulletin of Informatics and Cybernetics* (accepted).

## Appendix A

We prove Theorem 2.

PROOF. We use the Lindeberg's CLT; For example, see Feller (1966) for the details.

**Lemma 3** Suppose  $\{X_1, \dots, X_n\}$  is a sequence of independent random variables, each with the finite mean  $\mu_i$  and the finite variance  $\sigma_i^2$ . Put  $S_n^2 = \sum_{i=1}^n \sigma_i^2$ . Put  $S_n^2 := \sum_{i=1}^n \sigma_i^2$ , and let  $I_A$  denote indicator function. If, for any  $\varepsilon > 0$ , the Lindeberg's condition:

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n E[(X_i - \mu_i)^2 I_{\{|X_i - \mu_i| > \varepsilon S_n\}}] = 0 \quad (17)$$

is satisfied then, it holds that

$$\frac{1}{S_n} \sum_i (X_i - \mu_i) \xrightarrow{d} N(0, 1),$$

as  $n \rightarrow \infty$ .

From (2), we rewrite S-LLR as

$$\hat{m}(\theta; \kappa) = n^{-1} \mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta^T)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{Y}. \quad (18)$$

Put the vector  $\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta^T)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta = (c_1, \dots, c_n)$ , where  $c_i$  are any constants. Then, from (18) S-LLR is given by the average of  $c_i Y_i$ . That is,

$$\hat{m}(\theta; \kappa) = n^{-1} \sum_{i=1}^n c_i Y_i. \quad (19)$$

From combining (5) and (19), we obtain the sum of variances of  $c_i Y_i / \sqrt{R(K_\kappa)}$  is approximately equal to

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n \text{Var}_Y[c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n] \\ &= n^2 R(K_\kappa)^{-1} \text{Var}_Y[\hat{m}(\theta; \kappa) | \Theta_n] \\ &\simeq n^2 R(K_\kappa)^{-1} R(K_\kappa) \frac{v(\theta)}{nf(\theta)} \\ &= nv(\theta)/f(\theta). \end{aligned} \quad (20)$$

It follows from (20) that as  $n \rightarrow \infty$ ,  $S_n^2 \rightarrow \infty$ . If  $n$  is large enough, then  $E_Y[(Y_i - E_Y[Y_i])^2 I_{\{|Y_i - E_Y[Y_i]| > \varepsilon S_n\}} | \Theta_n]$  is equal to

$$\begin{aligned} &\lim_{n \rightarrow \infty} E_Y[(Y_i - E_Y[Y_i | \Theta_n])^2 I_{\{|Y_i - E_Y[Y_i | \Theta_n]| > \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y[Y_i | \Theta_n] \\ &\quad - \lim_{n \rightarrow \infty} E_Y[(Y_i - E_Y[Y_i | \Theta_n])^2 I_{\{|Y_i - E_Y[Y_i | \Theta_n]| \leq \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y[Y_i | \Theta_n] - \text{Var}_Y[Y_i | \Theta_n] \\ &= 0. \end{aligned} \quad (21)$$

From combining (20) and (21), it follows that

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n \mathbb{E}_Y [(c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n])^2 \mathbb{I}_{\{Y_i - \mathbb{E}_Y [Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\
&= \lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n c_i^2 R(K_\kappa)^{-1} \mathbb{E}_Y [(Y_i - \mathbb{E}_Y [Y_i | \Theta_n])^2 \mathbb{I}_{\{Y_i - \mathbb{E}_Y [Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\
&= 0.
\end{aligned} \tag{22}$$

From (22), we show that  $c_i Y_i / \sqrt{R(K_\kappa)}$  satisfies Lindeberg condition for any  $\varepsilon > 0$ . Therefore, from considering Lemma 3, (19), and (20), we obtain the following asymptotic distribution:

$$\begin{aligned}
& \frac{n}{\sqrt{n R(K_\kappa) v(\theta) / f(\theta)}} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \\
&= \frac{n}{\sqrt{n v(\theta) / f(\theta)}} [n^{-1} \sum_{i=1}^n \{c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n]\}] \\
&= \frac{1}{S_n} \sum_{i=1}^n \{c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n]\} \\
&\xrightarrow{d} N(0, 1),
\end{aligned} \tag{23}$$

as  $n \rightarrow \infty$ . Theorem 2 completes the proof from (23).  $\square$

## Appendix B

We prove Theorem 4.

PROOF. From Theorem 2 and Lemma 1, we obtain the following asymptotically normal distribution:

$$\sqrt{\frac{n}{\kappa^{1/2} / (2\pi^{1/2})}} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \xrightarrow{d} N(0, v(\theta) / f(\theta)). \tag{24}$$

Equation (24) is reduced to

$$n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \xrightarrow{d} N(0, v(\theta) / \{2\pi^{1/2} f(\theta)\}). \tag{25}$$

We obtain that  $n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - m(\theta)]$  is equal to

$$\begin{aligned}
n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - m(\theta)] &= n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n] \\
&\quad + n^{1/2} \kappa^{-1/4} \text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n].
\end{aligned} \tag{26}$$

We put  $\kappa = cn^\alpha$ . Then, recalling that the equation (9) gives that  $\text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n] = O(\kappa^{-1})$ , it follows that

$$\begin{aligned}
n^{1/2} \kappa^{-1/4} \text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n] &\propto n^{1/2} \kappa^{-5/4} \\
&= O_p(n^{(2-5\alpha)/4}).
\end{aligned} \tag{27}$$

From (27), we show that  $\alpha$  such as  $n^{(2-5\alpha)/4} = o_p(1)$  is  $\alpha > 2/5$ . Hence, if  $\alpha > 2/5$  and  $n \rightarrow \infty$ , then the second term of the right side in (26) is vanished. Therefore, from combining (25), and (26), it holds that

$$\begin{aligned} n^{1/2}\kappa^{-1/4}[\hat{m}(\theta; \kappa) - m(\theta)] &\simeq n^{1/2}\kappa^{-1/4}[\hat{m}(\theta; \kappa) - \mathbb{E}_Y[\hat{m}(\theta; \kappa)|\Theta_n]] \\ &\xrightarrow{d} N(0, v(\theta)/\{2\pi^{1/2}f(\theta)\}), \quad n \rightarrow \infty. \end{aligned} \quad (28)$$

Theorem 4 completes the proof from (28).  $\square$

## Appendix C

We prove Theorem 6

PROOF. From Theorem 2 and Lemma 2, we obtain the following asymptotically normal distribution:

$$(nh)^{1/2}[\hat{m}(\theta; h) - \mathbb{E}_Y[\hat{m}(\theta; h)|\Theta_n]] \xrightarrow{d} N(0, v(\theta)/\{\pi f(\theta)\}). \quad (29)$$

We show that  $(nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)]$  is equal to

$$\begin{aligned} (nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)] &= (nh)^{1/2}[\hat{m}(\theta; h) - \mathbb{E}_Y[\hat{m}(\theta; h)|\Theta_n] + \text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n]] \\ &= (nh)^{1/2}[\hat{m}(\theta; h) - \mathbb{E}_Y[\hat{m}(\theta; h)|\Theta_n] + (nh)^{1/2}\text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n]]. \end{aligned} \quad (30)$$

We put  $h = cn^\alpha$ . Then, recalling that equation (13) gives that  $\text{Bias}_Y[\hat{m}(\theta; \kappa)|\Theta_n] = O(h)$ , it follows that

$$\begin{aligned} (nh)^{1/2}\text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n] &\propto n^{1/2}h^{3/2} \\ &= O_p(n^{(1+3\alpha)/2}). \end{aligned} \quad (31)$$

From (31), we show that  $\alpha$  such as  $n^{(1+3\alpha)/2} = o_p(1)$  is  $\alpha < -1/3$ . Hence, if  $\alpha < -1/3$  and  $n \rightarrow \infty$ , then the second term of the right side in (30) is vanished. Therefore, from combining (29), and (30), it holds that

$$\begin{aligned} (nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)] &\simeq (nh)^{1/2}[\hat{m}(\theta; h) - \mathbb{E}_Y[\hat{m}(\theta; h)|\Theta_n]] \\ &\xrightarrow{d} N(0, v(\theta)/\{\pi f(\theta)\}), \quad n \rightarrow \infty. \end{aligned} \quad (32)$$

Theorem 6 completes the proof from (32).  $\square$

*Received October 13, 2017*

*Revised April 25, 2018*