

Discussion Paper

No. 28

Asymptotic Property of Wrapped Cauchy  
Kernel Density Estimation on the Circle

Yasuhito Tsuruta

Masahiko Sagae

# Asymptotic Property of Wrapped Cauchy Kernel Density Estimation on the Circle

Yasuhito Tsuruta \*      Masahiko Sagae †

2016/6/10

## 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.

## 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), C. C. Taylor (2008), Di Marzio et al. (2009, 2011).

Di Marzio et al. (2009, 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. (2009, 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

---

\*Graduate School of Human and Socio-Environment Studies, Kanazawa University

†School of Economics, Kanazawa University

line. However, the convergent rate of the AMISE can not be derived without specifying the kernel.

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

(a) it admits an uniformly convergent Fourier series,

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

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

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

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

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

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

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

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

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

**Definiton 3.** (*Kernel density estimator*)

Let  $\Theta_1, \dots, \Theta_n$  be random sample from the unknown circular density  $f(\theta)$ . Given a circular kernel  $K_\kappa$ , 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)  $\kappa$  increases as  $n \rightarrow \infty$ . for each  $j \in \mathbb{Z}^+$ ,  $\lim_{n \rightarrow \infty} \gamma_j(\kappa) = 1$ ,
- (B)  $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\kappa) = 0$ ,
- (C)  $f''$  is continuous and square-integrable,

the second sin-order kernel has,

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

where  $\int_0^{2\pi} \{g(\theta)\}^2 d\theta = R(g)$ ,  $\eta_2(K_\kappa) = \frac{1}{2}(1 - \gamma_2(\kappa))$  and  $R(K_\kappa) = (1 + 2 \sum_{i=1}^{\infty} \gamma_i^2(\kappa))/(2\pi)$ . 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  $K_\kappa$ . Di Marzio et al. (2009, 2011) derived AMISE of the VM kernel.

The VM kernel is defined as,

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

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,

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

**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 :

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

Under the condition that  $\kappa$  is sufficiently large, the asymptotic form (3) is reduced to,

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

We obtain the optimum  $\kappa^*$ :

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

From (4) and (5),  $AMISE_{VM} = O(n^{-4/5})$  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,

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

where  $\rho$  is concentration parameter. The characteristic function of the WC kernel is equal to,

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

**Theorem 2.** (AMISE of WC kernel)

The AMISE of the WC kernel is given as the follows,

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

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

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

See Appendix-A for the details.

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

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

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

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

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

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

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

**Theorem 3.** (*Asymptotic Normality*)

Let be  $\Theta_1, \Theta_2, \dots, \Theta_n \stackrel{i.i.d.}{\sim} f(\theta)$ ,  $h = cn^{-\alpha}$  and  $0 \leq h < 1$ . If  $\alpha > 1/3$ , hold, then

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

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 correspondence 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  $f_{\text{VM}}$ . The simplest plug-in rule uses  $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\kappa}))$  as the estimator for  $R(f'')$ , where  $\kappa$  is the concentration parameter of  $f_{\text{VM}}$  and  $\hat{\kappa}$  is the maximum likelihood estimator of  $\kappa$ .  $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\kappa}))$  corresponds to the following form:

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

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

- i) Estimate  $\hat{\kappa}$  from the sample and calculate  $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\kappa}))$  from  $\hat{\kappa}$ ,
- ii) WC kernel: Estimate  $\rho^*$  which minimize (6) by substituting  $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\kappa}))$  for  $R(f'')$  in (6),
- iii) VM kernel : Estimate  $\kappa^*$  by substituting  $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\kappa}))$  for  $R(f'')$  in (5).

Our simulation is given as,

### 1. The WC kernel:

- (a) Let a true density be a mixture of the von Mises density:

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

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  $\rho^*$  by the plug-in rule,
- (c) Let  $\text{ISE} = \int_0^{2\pi} \{\hat{f}(\theta; \rho) - f(\theta)\}^2 d\theta$ , Let  $\overline{\text{ISE}}(\hat{f}(\cdot; \rho))$  be the numerical integration of ISE. Calculate  $\overline{\text{ISE}}(\hat{f}(\cdot; \rho))$ ,
- (d) Repeat (a)-(c) 1000 times and compute  $\overline{\text{MISE}}(\hat{f}(\cdot; \rho)) = \sum_{i=1}^{1000} \overline{\text{ISE}}_i(\hat{f}(\cdot; \rho))/1000$ .

### 2. The VM kernel :

With the same procedures as (a) - (d) of the WC kernel, compute  $\overline{\text{MISE}}(\hat{f}(\cdot; \kappa^*))$ ,

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

|                | $n = 10$ | 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 |

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

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 multimodal and/or the heavy tail. This difference between the WC kernel and the VM kernel tends to become smaller as  $n$  become larger.

## 5 Conclusion

This paper shows that the WC kernel has the convergent rate  $O(n^{-2/3})$  of the AMISE and holds 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 result of our simulation shows that the WC kernel is superior to the VM kernel if a true density  $f$  has the multimodal and/or the heavy tail.



## 6 Appendix

### 6.1 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  $f(\theta + u) \simeq f(\theta) + f'(\theta) \sin(u) + f''(\theta) \sin^2(u)/2 + O(\sin^3(u))$ .

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

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

It follows from (15) that

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

□

## 6.2 Appendix-B

*proof of Theorem 3.* Put  $h = (1 - \rho^2)$ , Write (14) and (16) as,

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

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

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

The first term of (19) is equal to,

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

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

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

From (18), We obtain the following form:

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

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

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

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

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

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

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, then (24) is given as the following form:

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

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

## References

- [1] Davis, K.B. (1975). Mean Square Error Properties of Density Estimates. *The Annals of Statistics* **3**, 1025-1030.
- [2] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters* **79**, 2066-2075.
- [3] 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.
- [4] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. I (Third ed.)*. John Wiley & Sons, Inc. pp.244.
- [5] Garcia-Portugues, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics* **7**, 1655-1685.
- [6] Hall, P., Watson, G. S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, **74**, 751-762.
- [7] Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. John Wiley & Sons, Ltd, London.

- [8] Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley.
- [9] Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis* **52**, 3493-3500 .