No. 35

Asymptotic Properties of Circular Nonparametric Regression by applying Von Mises and Wrapped Cauchy Kernels

Tsuruta Yasuhito

Sagae Masahiko

9 June 2017

Asymptotic Properties for Circular Nonparametric Regression by von Mises and Wrapped Cauchy kernels

Yasuhito Tsuruta [*] Masahiko Sagae [†]

September 26, 2017

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.

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 Θ_i for $\theta_i \in [-\pi, \pi)$. Let the data set $\{(Y_1, \Theta_1), \ldots, (Y_n, \Theta_n)\}$ be i.i.d. Then, we consider that

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

where $v(\theta) =: \mathrm{Var}_Y[Y|\Theta = \theta]$ is the conditional variance, ε_i is a random variable on the real line with zero mean and unit variance, and a regression function $m(\theta) := \mathrm{E}_Y[Y|\Theta = \theta]$ 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 [1]. S-LLR $\hat{m}(\theta; \kappa)$ is defined as $\hat{\beta}_0$ that

[*]Graduate School of Human and Socio-Environment Studies, Kanazawa University
[†]School of Economics, Kanazawa University

minimizes

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

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

[1] 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 MISE. However, few study explored the theoretical properties for S-LLR employing another kernel functions such as 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 wrapped Cauchy (WC) kernels.

In section 2 we demonstrate the definitions of S-LLR and a class of kernels, and the MSE of S-LLR by [1]. 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 such as MISE and the asymptotic normality. And, we show 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 for Kernel density estimations for circular data, [5] 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

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

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

$$\boldsymbol{S_\theta} := \begin{bmatrix} 1 & \sin(\Theta_1 - \theta) \\ \vdots & \vdots \\ 1 & \sin(\Theta_n - \theta) \end{bmatrix}$$

is an $n \times 2$ design matrix, and $\boldsymbol{W_\theta} := \mathrm{diag}\{K_\kappa(\Theta_1 - \theta), \dots, K_\kappa(\Theta_n - \theta)\}$ is an $n \times n$ diagonal matrix.

We employ a class of kernels $K_\kappa(\theta)$ satisfying the following definition proposed by [1] and [2].

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:

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

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

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

We now define a jth sine-type moment as

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

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

$$\eta_2(K_\kappa) = (1 - \gamma_2(\kappa))/2. \tag{3}$$

Put $\boldsymbol{\Theta}_n := \{\Theta_1, \dots, \Theta_n\}$ and $R(g) := \int_{-\pi}^{\pi} g(\theta)^2 d\theta$. Then, let the conditional bias be $\mathrm{Bias}_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n] =: \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n] - m(\theta)$ and the conditional variance of S-LLR be $\mathrm{Var}_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n]$. [1] derived the following theorem regarding the bias and the variance.

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

i) $\lim_{n\to\infty} n^{-1} R(K_\kappa) = 0$.

ii) $\lim_{n\to\infty} \gamma_j(\kappa) = 1$.

iii) The marginal density $f(\theta)$ *is continuously differentiable, where* $f(\theta) > 0$ *for any* θ.

iv) The second derivative $m''(\theta)$ *and the conditional variance* $v(\theta)$ *are continuous, respectively.*

Then, the bias is approximately given by

$$Bias_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] \simeq \eta_2(K_\kappa)\frac{m''(\theta)}{2!}, \tag{4}$$

and the variance is approximately given by

$$Var_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] \simeq R(K_\kappa)\frac{v(\theta)}{nf(\theta)}. \tag{5}$$

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

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

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

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 of Theorem 1, it is needed to divide this two terms into κ and any constant part $C(K)$, but it is difficult problem to obtain kernel's general 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

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

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

$$\gamma_j(\kappa) = I_j(\kappa)/I_0(\kappa). \tag{6}$$

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 [4], the second sine-type moment for VM kernel is given by*

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

If κ is large enough, then from combining (7) *and (3.5.34) in [4], 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 [4], the term $R(K_\kappa)$ is given by

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

If κ is large enough, then from combining (8) *and (3.5.33) in [4], the term $R(K_\kappa)$ can approximate to*

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

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

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

$$Bias_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n] \simeq \frac{1}{2\kappa}m''(\theta), \tag{9}$$

and the variance is approximately given by

$$Var_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] \simeq \frac{\kappa^{1/2} v(\theta)}{2\pi^{1/2} n f(\theta)}. \tag{10}$$

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

$$AMISE_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] = \frac{1}{4\kappa^2}\int_{-\pi}^{\pi} m''(\theta)^2 f(\theta)d\theta + \frac{\kappa^{1/2}\int_{-\pi}^{\pi} v(\theta)d\theta}{2\pi^{1/2} n}. \tag{11}$$

The minimizer κ_* *of* (11) *is given by*

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

Therefore, the optimal $AMISE_Y[\hat{m}(\theta;\kappa_*)|\mathbf{\Theta}_n]$ *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* α *are any constants. Then, if* $\alpha > 2/5$ *and* $n \to \infty$, *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 ρ 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 $h = 1-\rho^2 \quad 0 < h < 1$. 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* $n \to \infty$*,* $h \to 0$ *and* $nh \to \infty$*. Then, the bias is approximately given by*

$$Bias_Y[\hat{m}(\theta;h)|\mathbf{\Theta}_n] \simeq h\frac{m''(\theta)}{4}, \tag{13}$$

and the variance is approximately given by

$$Var_Y[\hat{m}(\theta;h)|\mathbf{\Theta}_n]s \simeq (nh)^{-1}\frac{v(\theta)}{\pi f(\theta)}. \tag{14}$$

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

$$AMISE_Y[\hat{m}(\theta;h)|\mathbf{\Theta}_n] = \frac{h^2\int_{-\pi}^{\pi} m''(\theta)^2 f(\theta)d\theta}{16} + \frac{\int_{-\pi}^{\pi} v(\theta)d\theta}{\pi n h}. \tag{15}$$

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 $AMISE_Y[\hat{m}(\theta;h_*)|\mathbf{\Theta}_n]$ *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 a kernel employed. 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 Theorem 2 and Lemma 2.

Theorem 6. *Put $h = cn^{\alpha}$, where c and α are any constants. Then, if $\alpha < -1/3$ and $n \to \infty$, 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 theoretical aspects for VM kernel and WC kernel in the above sections. From practical view point we want to investigate the performances in finite samples for the both kernel through a numerical experiment.

We consider the following model:

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

where the regression function is $m(\theta) = 2+3\cos(\theta)+2\sin(3\theta)$, and Θ_i follow a circular uniform density: $f_{\mathrm{CU}}(\theta) = 1/(2\pi)$ for $\theta \in [-\pi, \pi)$, and are each independent.

We produce S-LLRs applying VM and WC kerneld with each the optimal parameter in the above settings, because we want to remove the influence of estimating the the optimal parameter from the estimator of $m(\theta)$. The optimal parameter of VM kernel is given by

$$\tilde{\kappa} = [333n/(2\pi^{1/2}t^2)]^{2/5}.$$

The optimal parameter of WC kernel is given by

$$\tilde{h} = [32t^2/(333n)]^{2/5}.$$

We should pay attention to the fact that the small sample properties for $\mathrm{MISE}_Y[\hat{m}(\theta; \cdot)|\mathbf{\Theta}_n]$ are greatly depend on sampling explanatory variables $\mathbf{\Theta}_n$. Therefore, sampling 100 samples of $\mathbf{\Theta}_n$, we calculate the average of MISE: Ave.MISE $= \sum_{j=1}^{100} \mathrm{MISE}_{j,Y}[\hat{m}(\theta; \cdot)|\mathbf{\Theta}_n]/100$.

The numerical experiment executes the following seven procedures:

1. Generate a random sample $\{\Theta_1, \dots, \Theta_n\}$ distributed as the circular uniform density $f_{\mathrm{CU}}(\theta)$.

2. Generate a random sample $\{\varepsilon_1, \dots, \varepsilon_n\}$ distributed as the normal distribution $\mathrm{N}(0, t^2)$.

3. Generate a random sample $\{Y_1, \dots, Y_n\}$ from 1–2.

4. Give VM kernel $\hat{m}(\theta; \tilde{\kappa})$ and WC kernel $\hat{m}(\theta; \tilde{h})$, respectively.
5. Calculate the numerical integrals $\mathrm{ISE_{VM}} := \int_{-\pi}^{\pi} \{\hat{m}(\theta; \tilde{\kappa}) - m(\theta)\}^2/(2\pi) d\theta$ and $\mathrm{ISE_{WC}} := \int_{-\pi}^{\pi} \{\hat{m}(\theta; \tilde{\rho}) - m(\theta)\}^2/(2\pi) d\theta$, respectively.
6. Repeat 1000 times from 2 to 5, and calculate $\mathrm{MISE_{VM}} := \sum_j \mathrm{ISE}_{\mathrm{VM},j}/1000$ and $\mathrm{MISE_{WC}} := \sum_j \mathrm{ISE}_{\mathrm{WC},j}/1000$, respectively.
7. Repeat 100 times from 1 to 6, and calculate $\mathrm{Ave.MISE_{VM}} := \sum_j \mathrm{MISE}_{\mathrm{VM},j}/100$ and $\mathrm{Ave.MISE_{WC}} := \sum_j \mathrm{MISE}_{\mathrm{WC},j}/100$, respectively.

Table 1 and Table 2 show that WC kernel outperforms VM kernel when $n \leq 20$, but VM kernel well performs than WC kernel when $n \geq 30$. Table 3 and Table 4 indicate that WC kernel's standard deviations with respect to the MISE are smaller than VM kernel's standard deviations. In other words, WC kernel is stable estimator than VM kernel in the small samples. These results show the advantage's for WC kernel in the small samples. The better performances of VM kernel when $n \geq 30$ correspond to on the large sample properties for VM and WC kernels.

6 Conclusion

We have shown the theoretical properties for S-LLRs applying VM and WC kernels that involve the asymptotic normality and the MISE. The convergence rate of the MISE of VM kernel is $O_p(n^{-4/5})$, but the convergence rate of that of WC kernel is $O_p(n^{-2/3})$. Our numerical experiment have indicated that WC kernel outperforms VM kernel in small samples, but in large samples VM kernel well performs than WC kernel.

Table 1: The values are the averages of the weighted conditional MISE for VM kernel: $\mathrm{Ave.MISE_{VM}} := \sum_j \mathrm{MISE}_{\mathrm{VM},j}/100$. n are the sample sizes, and t are the standard devision of errors.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	200.667	3.938	0.363	0.246	0.132	0.057
$t = 1.0$	11.314	1.441	0.678	0.401	0.315	0.158
$t = 1.5$	7.575	2.008	0.987	0.733	0.566	0.291
$t = 2.0$	12.122	2.805	1.470	1.076	0.840	0.447

Table 2: The values are the averages of the weighted conditional MISE for WC kernel: Ave.$\mathrm{MISE}_{\mathrm{WC}} := \sum_j \mathrm{MISE}_{\mathrm{WC},j}/100$. n are the sample sizes, and t are the standard devision of errors.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	**2.090**	**1.075**	0.594	0.421	0.319	0.148
$t = 1.0$	**2.627**	**1.258**	0.846	0.597	0.519	0.292
$t = 1.5$	**2.713**	**1.650**	1.165	0.938	0.815	0.476
$t = 2.0$	**4.161**	**2.256**	1.602	1.254	1.061	0.657

Table 3: The values are the standard devisions of the weighted conditional MISE for VM kernel. $n = 10, 20, 30, 40, 50, 100$ are the sample sizes, and $t = 0.5, 1.0, 1.5, 2.0$ are the standard devision of errors.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	1784.428	28.477	0.390	0.377	0.040	0.009
$t = 1.0$	27.064	1.708	0.473	0.086	0.055	0.012
$t = 1.5$	18.149	2.043	0.407	0.303	0.086	0.021
$t = 2.0$	31.531	2.960	0.405	0.226	0.112	0.026

Table 4: The values are the standard devisions of the weighted conditional MISE for WC kernel. $n = 10, 20, 30, 40, 50, 100$ are the sample sizes, and $t = 0.5, 1.0, 1.5, 2.0$ are the standard devision of errors.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	**1.197**	**0.814**	0.395	0.306	0.137	0.051
$t = 1.0$	**1.267**	**0.648**	**0.359**	0.154	0.155	0.046
$t = 1.5$	**1.080**	**0.593**	**0.325**	0.228	0.228	0.063
$t = 2.0$	**3.093**	**0.848**	**0.345**	0.226	0.143	0.073

7 Acknowledgements

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

References

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

[2] 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.

[3] Feller, M. (1966). *An Introduction to Probability Theory and Its Applications. II (Second ed.)*, John Wiley & Sons, Inc.

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

[5] Tsuruta, Y., and Sagae, M. (2017). Asymptotic Property of Wrapped Cauchy Kernel Density Estimation on the Circle. *Bulletin of Informatics and Cybernetics* (accepted).

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

Appendix A

We prove Theorem 2.

Proof. We use the Lindeberg's CLT; For example, see [3] for the details.

Lemma 3. *Suppose* $\{X_1, \dots X_n\}$ *is a sequence of independent random variables, each with the finite mean* μ_i *and the finite variance* σ_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:*

$$\lim_{n\to\infty} \frac{1}{S_n^2} \sum_{i=1}^n E[(X_i - \mu_i)^2 I_{\{|X_i-\mu_i|>\varepsilon S_n\}}] = 0 \tag{16}$$

is satisfied then, it holds that

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

as $n\to\infty$.

From (2), we rewrite S-LLR as

$$\hat{m}(\theta;\kappa) = n^{-1}\boldsymbol{e}_1^T(n^{-1}\boldsymbol{S}_{\boldsymbol{\theta}}^T\boldsymbol{W}_{\boldsymbol{\theta}}\boldsymbol{S}_{\boldsymbol{\theta}}^T)^{-1}\boldsymbol{S}_{\boldsymbol{\theta}}^T\boldsymbol{W}_{\boldsymbol{\theta}}\boldsymbol{Y}. \tag{17}$$

Put the vector $\boldsymbol{e}_1^T(n^{-1}\boldsymbol{S}_{\boldsymbol{\theta}}^T\boldsymbol{W}_{\boldsymbol{\theta}}\boldsymbol{S}_{\boldsymbol{\theta}})^{-1}\boldsymbol{S}_{\boldsymbol{\theta}}^T\boldsymbol{W}_{\boldsymbol{\theta}} = (c_1,\dots,c_n)$, where c_i are any constants. Then, from (17) S-LLR is given by the average of c_iY_i. That is,

$$\hat{m}(\theta;\kappa) = n^{-1}\sum_{i=1}^n c_iY_i. \tag{18}$$

From combining (5) and (18), we obtain the sum of variances of $c_iY_i/\sqrt{R(K_\kappa)}$ is approximately equal to

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n \mathrm{Var}_Y[c_iY_i/\sqrt{R(K_\kappa)}|\boldsymbol{\Theta}_n] \\ &= n^2R(K_\kappa)^{-1}\mathrm{Var}_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n] \\ &\simeq n^2R(K_\kappa)^{-1}R(K_\kappa)\frac{v(\theta)}{nf(\theta)} \\ &= nv(\theta)/f(\theta). \end{aligned} \tag{19}$$

It follows from (19) that as $n\to\infty$, $S_n^2\to\infty$. If n is large enough, then $\mathrm{E}_Y[(Y_i-\mathrm{E}_Y[Y_i])^2\mathrm{I}_{\{(Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n])>\varepsilon S_n\}}|\boldsymbol{\Theta}_n]$ is equal to

$$\begin{aligned} &\lim_{n\to\infty}\mathrm{E}_Y[(Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n])^2\mathrm{I}_{\{Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n]>\varepsilon S_n\}}|\boldsymbol{\Theta}_n] \\ &\quad= \mathrm{Var}_Y[Y_i|\boldsymbol{\Theta}_n] \\ &\quad\ - \lim_{n\to\infty}\mathrm{E}_Y[(Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n])^2\mathrm{I}_{\{Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n]\le\varepsilon S_n\}}|\boldsymbol{\Theta}_n] \\ &\quad= \mathrm{Var}_Y[Y_i|\boldsymbol{\Theta}_n]-\mathrm{Var}_Y[Y_i|\boldsymbol{\Theta}_n] \\ &\quad= 0. \end{aligned} \tag{20}$$

From combining (19) and (20), it follows that

$$\begin{aligned} &\lim_{n\to\infty}\frac{1}{S_n^2}\sum_{i=1}^n \mathrm{E}_Y[(c_iY_i/\sqrt{R(K_\kappa)}-\mathrm{E}_Y[c_iY_i/\sqrt{R(K_\kappa)}|\boldsymbol{\Theta}_n])^2\mathrm{I}_{\{Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n]>\varepsilon S_n\}}|\boldsymbol{\Theta}_n] \\ &\quad= \lim_{n\to\infty}\frac{1}{S_n^2}\sum_{i=1}^n c_i^2R(K_\kappa)^{-1}\mathrm{E}_Y[(Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n])^2\mathrm{I}_{\{Y_i-\mathrm{E}_Y[Y_i|\boldsymbol{\Theta}_n]>\varepsilon S_n\}}|\boldsymbol{\Theta}_n] \\ &\quad= 0. \end{aligned} \tag{21}$$

From (21), we show that $c_iY_i/\sqrt{R(K_\kappa)}$ satisfies Linderberg condition for any $\varepsilon > 0$. Therefore, from considering Lemma 3, (18), and (19), we obtain the following asymptotic distribution:

$$\begin{aligned}
&\frac{n}{\sqrt{nR(K_\kappa)v(\theta)/f(\theta)}}[\hat{m}(\theta;\kappa) - \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n]] \\
&\qquad = \frac{n}{\sqrt{nv(\theta)/f(\theta)}}[n^{-1}\sum_{i=1}^{n}\{c_iY_i/\sqrt{R(K_\kappa)} - \mathrm{E}_Y[c_iY_i/\sqrt{R(K_\kappa)}|\mathbf{\Theta}_n]\}] \\
&\qquad = \frac{1}{S_n}\sum_{i=1}^{n}\{c_iY_i/\sqrt{R(K_\kappa)} - \mathrm{E}_Y[c_iY_i/\sqrt{R(K_\kappa)}|\mathbf{\Theta}_n]\} \\
&\qquad \xrightarrow{d} \mathrm{N}(0,1), \qquad (22)
\end{aligned}$$

as $n \to \infty$. Theorem 2 completes the proof from (22). □ □

Appendix B

We prove Theorem 4.

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

$$\sqrt{\frac{n}{\kappa^{1/2}/(2\pi^{1/2})}}[\hat{m}(\theta;\kappa) - \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n]] \xrightarrow{d} \mathrm{N}(0, v(\theta)/f(\theta)). \qquad (23)$$

Equation (27) is reduced to

$$n^{1/2}\kappa^{-1/4}[\hat{m}(\theta;\kappa) - \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n]] \xrightarrow{d} \mathrm{N}(0, v(\theta)/\{2\pi^{1/2}f(\theta)\}). \qquad (24)$$

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

$$\begin{aligned}
n^{1/2}\kappa^{-1/4}[\hat{m}(\theta;\kappa) - m(\theta)] &= n^{1/2}\kappa^{-1/4}[\hat{m}(\theta;\kappa) - \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] \\
&\quad + n^{1/2}\kappa^{-1/4}\mathrm{Bias}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n]. \qquad (25)
\end{aligned}$$

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

$$\begin{aligned}
n^{1/2}\kappa^{-1/4}\mathrm{Bias}_Y[\hat{m}(\theta;\kappa)|\mathbf{\Theta}_n] &\propto n^{1/2}\kappa^{-5/4} \\
&= O_p(n^{(2-5\alpha)/4}). \qquad (26)
\end{aligned}$$

From (26), we show that α such as $n^{(2-5\alpha)/4} = o_p(1)$ is $\alpha > 2/5$. Hence, if $\alpha > 2/5$ and $n \to \infty$, then the second term of the right side in (25) is vanished. Therefore, from combining (24), and (25), it holds that

$$n^{1/2}\kappa^{-1/4}[\hat{m}(\theta;\kappa) - m(\theta)] \simeq n^{1/2}\kappa^{-1/4}[\hat{m}(\theta;\kappa) - \mathrm{E}_Y[\hat{m}(\theta;\kappa)|\boldsymbol{\Theta}_n]] \xrightarrow{d} \mathrm{N}(0, v(\theta)/\{2\pi^{1/2}f(\theta)\}) \quad n \to \infty. \tag{27}$$

Theorem 4 completes the proof from (27). □ □

Appendix C

We prove Theorem 6

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

$$(nh)^{1/2}[\hat{m}(\theta;h) - \mathrm{E}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n]] \xrightarrow{d} \mathrm{N}(0, v(\theta)/\{\pi f(\theta)\}). \tag{28}$$

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

$$\begin{aligned}(nh)^{1/2}[\hat{m}(\theta;h) - m(\theta)] &= (nh)^{1/2}[\hat{m}(\theta;h) - \mathrm{E}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n] + \mathrm{Bias}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n]] \\ &= (nh)^{1/2}[\hat{m}(\theta;h) - \mathrm{E}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n] + (nh)^{1/2}\mathrm{Bias}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n].\end{aligned} \tag{29}$$

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

$$\begin{aligned}(nh)^{1/2}\mathrm{Bias}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n] &\propto n^{1/2}h^{3/2} \\ &= O_p(n^{(1+3\alpha)/2})\end{aligned} \tag{30}$$

From (30), we show that α such as $n^{(1+3\alpha)/2} = o_p(1)$ is $\alpha < -1/3$. Hence, if $\alpha < -1/3$ and $n \to \infty$, then the second term of the right side in (29) is vanished. Therefore, from combining (28), and (29), it holds that

$$(nh)^{1/2}[\hat{m}(\theta;h) - m(\theta)] \simeq (nh)^{1/2}[\hat{m}(\theta;h) - \mathrm{E}_Y[\hat{m}(\theta;h)|\boldsymbol{\Theta}_n]] \xrightarrow{d} \mathrm{N}(0, v(\theta)/\{\pi f(\theta)\}), \quad n \to \infty. \tag{31}$$

Theorem 6 completes the proof from (31). □ □