A CLASS OF NONPARAMETRIC DENSITY DERIVATIVE ESTIMATORS BASED ON GLOBAL LIPSCHITZ CONDITIONS

Kairat Mynbaev[a], Carlos Martins-Filho[b,c] and Aziza Aipenova[d,e]

[a]*International School of Economics, Kazakh-British Technical University, Almaty, Kazakhstan*
[b]*Department of Economics, University of Colorado at Boulder, Boulder, CO, USA*
[c]*IFPRI, Washington, DC, USA*
[d]*Department of Mechanics and Mathematics, Kazakh National University, Almaty, Kazakhstan*
[e]*Faculty of Engineering and Natural Sciences, Suleyman Demirel University, Almaty, Kazakhstan*

ABSTRACT

Estimators for derivatives associated with a density function can be useful in identifying its modes and inflection points. In addition, these estimators play an important role in plug-in methods associated with bandwidth selection in nonparametric kernel density estimation. In this paper, we extend

Essays in Honor of Aman Ullah
Advances in Econometrics, Volume 36, 591–615

ISSN: 0731-9053/doi:10.1108/S0731-905320160000036026

the nonparametric class of density estimators proposed by Mynbaev and Martins-Filho (2010) to the estimation of m*-order density derivatives. Contrary to some existing derivative estimators, the estimators in our proposed class have a full asymptotic characterization, including uniform consistency and asymptotic normality. An expression for the bandwidth that minimizes an asymptotic approximation for the estimators' integrated squared error is provided. A Monte Carlo study sheds light on the finite sample performance of our estimators and contrasts it with that of density derivative estimators based on the classical Rosenblatt–Parzen approach.*

Keywords: Nonparametric derivative estimation; Lipschitz conditions

AMS-MS classifications: 62G07; 62G20

JEL classifications: C14; C18

1. INTRODUCTION

Let f be a density associated with a real random variable X and $\{X_j\}_{j=1}^n$ be an independent and identically distributed random sample of size n from f. The Rosenblatt–Parzen estimator for the density f evaluated at $x \in \mathbb{R}$ is given by

$$\hat{f}_{RP}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n}K\left(\frac{x - X_j}{h_n}\right),$$

where $h_n > 0$ is a global bandwidth and K is a kernel on $\mathbb{R}$ satisfying

$$\int_{-\infty}^{+\infty} K(t)dt = 1. \tag{1}$$

If f and K are $m \in \mathbb{N}$ times continuously differentiable, with $f^{(m)}$ and $K^{(m)}$ denoting their mth order derivatives, the most commonly used estimator of $f^{(m)}$ at $x \in \mathbb{R}$ (Bhattacharya, 1967) is given by $\hat{f}_{RP}^{(m)}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^{m+1}}K^{(m)}\left(\frac{x-X_j}{h_n}\right)$.

Estimators for $f^{(m)}$ are important in various contexts. They can be used to evaluate the location of modes and inflection points of f, to construct plug-in bandwidths for kernel density estimation of f, and can be applied to the estimation of scores in certain additive models (Härdle & Stoker, 1989). The asymptotic properties of $\hat{f}_{RP}^{(m)}(x)$ have been studied by, among others, Bhattacharya (1967), Schuster (1969), and Silverman (1978). Singh (1977, 1979, 1987) shows that it is possible to reduce bias and improve the mean integrated squared error (MISE) of $\hat{f}_{RP}^{(m)}(x)$ by considering restrictions on the class of kernels K used in its construction. Similar efforts have been undertaken by Muller (1984) and Henderson and Parmeter (2012).

Recently, Mynbaev and Martins-Filho (2010) proposed a class of nonparametric density estimators that attains bias reduction relative to $\hat{f}_{RP}(x)$ by imposing global higher order Lipschitz conditions on f. Usually, the order of the bias for $\hat{f}_{RP}(x)$ is established by requiring that f be r-times ($r \in \mathbb{N}$) differentiable. They show that r-times differentiability is stronger than a Lipschitz order r. Hence, although some smoothness is still required to attain a suitable order for the bias, the constraint on the class of densities containing f is milder than what is traditionally required. In practice, certain discontinuous densities satisfy global Lipschitz conditions of a certain order, but are not differentiable of the same order.[1]

In this paper, we propose a new class of estimators for $f^{(m)}$ by considering m-order derivatives of the kernel density estimators in the class proposed in Mynbaev and Martins-Filho (2010). We provide a full asymptotic characterization of the new density derivative estimators, including uniform consistency, asymptotic normality and give exact convergence rates. An important by-product of our results is an expression and the exact order for the bias for the density estimators proposed in Mynbaev and Martins-Filho (2010). There, they only provide the order of the bound on the absolute bias. This is useful, since it allows for our discussion of optimal bandwidth selection based on the minimization of an asymptotic approximation for the integrated mean-squared error.

Besides this introduction, this paper contains four more sections. The next section provides new estimators for $f^{(m)}$ based on a class of density estimators proposed in Mynbaev and Martins-Filho (2010) and a fundamental integral representation for their bias. The following section provides asymptotic properties of our estimators and discusses optimal bandwidth selection. Then, we provide a small Monte Carlo study that gives some evidence on the small sample properties of our estimators and compares their performance to that of $\hat{f}_{RP}^{(m)}$. Finally, the last section provides a conclusion. All proofs and technical lemmas are collected in the appendix.

2. A CLASS OF ESTIMATORS FOR $f^{(m)}(x)$ AND THEIR BIAS

The properties of nonparametric density estimators are traditionally obtained by imposing smoothness conditions on the underlying density f. Smoothness can be regulated by finite differences, which can be defined as forward, backward, or centered. The corresponding examples of finite first-order differences for a function $f(x)$ are $f(x+h)-f(x)$, $f(x)-f(x-h)$, and $f(x+h)-f(x-h)$, where $h \in \mathbb{R}$. We will focus on centered even-order differences because, as will soon become apparent, the resulting kernels are symmetric. Let $C_{2k}^l = \frac{(2k)!}{(2k-l)!l!}$, $l=0,\ldots,2k,\ k\in\mathbb{N}$ be the binomial coefficients, $c_{k,s}=(-1)^{s+k}C_{2k}^{s+k},\ s=-k,\ldots,k$ and

$$\Delta_h^{2k} f(x) = \sum_{s=-k}^{k} c_{k,s} f(x+sh), \quad h \in \mathbb{R}. \tag{2}$$

For example, taking $k=1,2$ we have $\Delta_h^2 f(x) = [f(x+h)-f(x)]-[f(x)-f(x-h)]$ and $\Delta_h^4 f(x) = [f(x+2h)-f(x+h)]-3[f(x+h)-f(x)]+3[f(x)-f(x-h)]-[f(x-h)-f(x-2h)]$.

We say that a function $f(x):\mathbb{R}\to\mathbb{R}$ satisfies the Lipschitz condition of order $2k$ if for any $x\in\mathbb{R}$ there exist $H(x)>0$ and $\varepsilon(x)>0$ such that $\left|\Delta_h^{2k} f(x)\right| \le H(x)h^{2k}$ for all h such that $|h|\le\varepsilon(x)$. We call $H(x)$ a Lipschitz constant and $\varepsilon(x)$ a Lipschitz radius.[2] For a kernel K, Mynbaev and Martins-Filho (2010) define a class of kernels $\{M_k(x)\}_{k=1,2,3,\ldots}$ where

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right). \tag{3}$$

K is called a seed kernel for M_k. The main impetus for the definition of $M_k(x)$ is that it allows us to express the bias of our proposed estimator

$$\hat{f}_k(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h_n} M_k\left(\frac{x-X_j}{h_n}\right) \quad \text{for } k=1,2,\ldots$$

in terms of centered even-order differences of $f(x)$. Let $\lambda_{k,s} = \frac{(-1)^{s+1}(k!)^2}{(k+s)!(k-s)!}$, $s=1,\ldots,k$ and since $-\frac{c_{k,s}}{c_{k,0}} = -\frac{c_{k,-s}}{c_{k,0}} = \lambda_{k,s},\ s=1,\ldots,k$, (3) can also be written as $M_k(x) = \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s}\left(K\left(\frac{x}{s}\right)+K\left(-\frac{x}{s}\right)\right)$. It follows by construction that M_k is symmetric, that is, $M_k(x)=M_k(-x),\ x\in\mathbb{R}$. Since the coefficients $c_{k,s}$ satisfy $\sum_{|s|=0}^{k} c_{k,s} = (1-1)^{2k} = 0$, we have $-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k} c_{k,s} = 1$ or $\sum_{s=1}^{k}\lambda_{k,s} = \frac{1}{2}$.

Consequently, Eqs. (1) and (3) imply that $\int_{-\infty}^{+\infty} M_k(x)dx = \sum_{s=1}^{k}\frac{\lambda_{k,s}}{s}\left(\int_{-\infty}^{+\infty} K\left(\frac{x}{s}\right) dx + \int_{-\infty}^{+\infty} K\left(-\frac{x}{s}\right)dx\right) = 1$, establishing that $\{M_k(x)\}_{k=1,2,\ldots}$ is a class of kernels. Tsybakov (2009) provides several choices for a seed kernel K, but perhaps the most popular would be a Gaussian density. In this case, $\hat{f}_k(x)$ has derivatives of all orders. It should also be noted that when K is symmetric $\hat{f}_1(x)$ is the traditional Rosenblatt–Parzen density estimator. We define a new class of $m=1,2,\ldots$ order nonparametric estimators for $f^{(m)}$ by

$$\hat{f}_k^{(m)}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^{m+1}}M_k^{(m)}\left(\frac{x-X_j}{h_n}\right) = \frac{1}{n}\sum_{j=1}^{n}u_j \tag{4}$$

where $u_j = \frac{1}{h_n^{m+1}}M_k^{(m)}\left(\frac{x-X_j}{h_n}\right)$ and

$$M_k^{(m)}(x) = -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|s^m}K^{(m)}\left(\frac{x}{s}\right). \tag{5}$$

It follows from the fact that $\{X_j\}_{j=1}^n$ is an independent and identically distributed random sample that

$$E\left(\hat{f}_k^{(m)}(x)\right) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^{m+1}}E\left(M_k^{(m)}\left(\frac{x-X_j}{h_n}\right)\right) = \frac{1}{n}\sum_{j=1}^{n}E(u_j) = E(u_1) \tag{6}$$

and

$$\begin{aligned} V\left(\hat{f}_k^{(m)}(x)\right) &= V\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^{m+1}}M_k^{(m)}\left(\frac{x-X_j}{h_n}\right)\right) = \frac{1}{n^2}\sum_{j=1}^{n}V(u_j) \\ &= \frac{V(u_1)}{n} = \frac{1}{n}(E(u_1^2) - E(u_1)^2). \end{aligned} \tag{7}$$

As in the existing literature, restrictions on K and f are needed to obtain a suitable representation for the bias and variance of the density derivative estimators. Hence, we assume that

Assumption 1.

(a) *K is symmetric and belongs to the weighted Sobolev space with norm $\|K\|_{W_1^m} = \int_{\mathbb{R}}\left(|K(t)| + |K^{(m)}(t)|\right)|t|dt$.*

(b) $\max\left\{|f(s)|, \ldots, |f^{(m-1)}(s)|\right\} = O(s), |s| \to \infty$.

Assumption 1 is used to obtain an integral representation for the bias $B(\hat{f}_k^{(m)}(x)) = E(\hat{f}_k^{(m)}(x)) - f^{(m)}(x)$ of $\hat{f}_k^{(m)}(x)$ in terms of centered even order differences of $f^{(m)}(x)$. Most of other results depend on this representation.

Theorem 1. *Under Assumption 1, for any $h_n > 0$, $B(\hat{f}_k^{(m)}(x)) = -\frac{1}{c_{k,0}}\int_{-\infty}^{+\infty}K(t)\Delta_{h_n t}^{2k}f^{(m)}(x)dt$.*

3. ASYMPTOTIC CHARACTERIZATION OF $\hat{f}_k^{(m)}(x)$

3.1. Uniform Consistency and Orders for Bias and Variance

We start our investigation of the asymptotic behavior of $\hat{f}_k^{(m)}(x)$ by providing conditions under which the estimator is asymptotically uniformly unbiased and uniformly consistent. To establish the uniform consistency $\hat{f}_k^{(m)}(x)$, we make the following assumption:

Assumption 2.
(a) The characteristic function ϕ_K of K satisfies $\int_{\mathbb{R}} |s^m \phi_K(s)|ds < \infty$; (b) $f^{(m)}(x)$ is bounded and uniformly continuous in $\mathbb{R}$; (c) $nh_n^{2m+2} \to \infty$ as $n \to \infty$.

Theorem 2. *Suppose that Assumption 2(a) and (c) hold. Then,*

$$\lim_{n\to\infty} E\left(\sup_{x\in\mathbb{R}} |\hat{f}_k^{(m)}(x) - E(\hat{f}_k^{(m)}(x))|\right) = 0.$$

Let also Assumption 2(b) hold. Then $\hat{f}_k^{(m)}(x)$ is uniformly consistent.

We note that the rate of decay of the bandwidth in Assumption 2(c), needed for the uniform consistency of $\hat{f}_k^{(m)}(x)$, could potentially be relaxed by, for example, limiting uniform consistency to restricted (compact) subsets of $\mathbb{R}$. In the following theorem, we provide exact orders for the bias and variance of the estimators. As a result, it can be promptly verified that the usual rate of decay of h_n implied by $nh_n^{2m+1} \to \infty$ as $n \to \infty$ is sufficient for pointwise consistency. The theorem depends on the following assumption.

Assumption 3.
(a) $f^{(m)}(x)$ is bounded and continuous in $\mathbb{R}$; (b) there exist functions $H_{2k,m}(x) > 0$ and $\varepsilon_{2k}(x) > 0$ such that

$$\left|\Delta_h^{2k} f^{(m)}(x)\right| \le H_{2k,m}(x)h^{2k} \quad \text{for all } |h| \le \varepsilon_{2k}(x); \tag{8}$$

(c) $\int_{-\infty}^{\infty} |K(t)|t^{2k}dt < \infty$.

Theorem 3. *Suppose that Assumptions 1 and 3 hold. Then, for all $x \in \mathbb{R}$ and $0 < h_n \le \varepsilon_{2k}(x)$*

$$\left|B(\hat{f}_k^{(m)}(x))\right| \le ch_n^{2k}\left(H_{2k,m}(x) + \varepsilon_{2k}^{-2k}(x)\right) \tag{9}$$

where the constant c does not depend on x or h_n. Suppose additionally that $\int_{-\infty}^{\infty}|K^{(m)}(t)|^2(1+|t|)dt<\infty$ and there exist functions $H_1(x)>0$ and $\varepsilon_1(x)>0$ such that

$$|f(x)-f(x-h)|\le H_1(x)|h| \quad \text{for all } |h|\le\varepsilon_1(x). \tag{10}$$

Then, for all $x\in\mathbb{R}$ and $0<h_n\le\min\{\varepsilon_{2k}(x),\varepsilon_1(x)\}$

$$\begin{aligned}V(\hat{f}_k^{(m)}(x)) &= \frac{1}{nh_n^{2m+1}}\Big\{f(x)\int_{-\infty}^{\infty}\Big(M_k^{(m)}(t)\Big)^2dt+R_1(x,h_n)\\&\quad -h_n[f^{(m)}(x)+R_{2k}(x,h_n,m)]^2\Big\},\end{aligned} \tag{11}$$

where the residuals satisfy

$$|R_1(x,h_n)|\le c_1|h_n|(H_1(x)+\varepsilon_1^{-1}(x)),\quad |R_{2k}(x,h_n,m)|\le c_2h_n^{2k}(H_{2k,m}(x)+\varepsilon_{2k}^{-2k}(x)) \tag{12}$$

with constants c_1 and c_2 independent of x and h_n.

3.2. Integrated Mean-Squared Error and Bandwidth Choice

We consider optimal choice of bandwidth by minimizing the Integrated Mean Squared Error (IMSE),

$$IMSE(\hat{f}^{(m)}(x))=\int_{\mathbb{R}}\Big(V(\hat{f}^{(m)}(x))+B(\hat{f}^{(m)}(x))^2\Big)dx.$$

The precise value of IMSE, as a function of h_n, is usually impossible to obtain. The common approach is to derive asymptotic approximations of variance and bias, as $h_n\to 0$, and plug those approximations into IMSE to obtain an approximation of type $IMSE\asymp\varphi(h_n)$ where φ depends on h_n, n and some well-defined constants. Then minimization of φ over h_n yields an expression of the optimal h_n as a function of the sample size. This is the approach we take up here. The result we formulate below, when $m=0$, is better than Theorem 8 in Mynbaev and Martins-Filho (2010). In the latter theorem, IMSE is bounded above by $\varphi(h_n)$, while here we obtain the asymptotic expression for IMSE. Our results depend on two auxiliary lemmas that are given in the appendix.

In the next theorem, we derive the exact order of bias, as $h_n\to 0$. The result is stronger than the upper bound Eq. (9) and, correspondingly, it requires stronger conditions. The result is also new for $m=0$ as Mynbaev and Martins-Filho (2010) did not derive the exact order of bias. We need the following assumption.

Assumption 4.
(a) $\int |K(t)||t|^{2k+1}dt = \beta_{2k+1}(K) < \infty$; (b) $f^{(m+2k)}$ is absolutely continuous, bounded and satisfies the following Lipschitz condition: there exist positive functions $H_{1,m}(x), \varepsilon(x)$ such that

$$\left| f^{(m+2k)}(x) - f^{(m+2k)}(x-h) \right| \leq H_{1,m}(x)|h| \quad \text{for all } |h| \leq \varepsilon(x). \tag{13}$$

Theorem 4. *Suppose that Assumptions 1 and 4 hold. Then*

$$B(\hat{f}^{(m)}(x)) = cf^{(m+2k)}(x)h_n^{2k} + O(h_n^{2k+1}G(x)) \tag{14}$$

where $c = (-1)^{k+1}\alpha_{2k}(K)\frac{(k!)^2}{(2k)!}$, $\alpha_{2k}(K) = \int_{\mathbb{R}} K(t)t^{2k}dt$ and $G(x) = H_{1,m}(x) + \frac{2\|f^{(m+2k)}\|_C}{\varepsilon(x)}$.

We note that if the Lipschitz condition in Eq. (13) is uniform on $\mathbb{R}$, then we can write $B(\hat{f}^{(m)}(x)) = cf^{(m+2k)}(x)h_n^{2k} + O(h_n^{2k+1})$. In the next theorem, we obtain the optimal bandwidth by minimizing the asymptotic expression for IMSE.

Theorem 5. *In addition to the conditions imposed in Theorem 3 for validity of Eq. (11), let us assume the conditions $f^{(m)}$, H_1, $\varepsilon_1^{-1} \in L_1$, H_{2k}, $\varepsilon_{2k}^{-2k} \in L_2$ that provide integrability in x of the right side of Eq. (11). In addition to the conditions of Theorem 4, suppose that $f^{(m+2k)}$, H, $\frac{1}{\varepsilon} \in L_2$ which makes sure that the right side of Eq. (14) is square integrable. Then*

$$IMSE = h_n^{4k}\left[c^2\|f^{(m+2k)}\|_{L_2}^2 + O(h_n)\right] + \frac{1}{nh_n^{2m+1}}\left[\|M_k^{(m)}\|_{L_2}^2 + O(h_n)\right]. \tag{15}$$

where c is the constant from Theorem 4. Hence, the function $\varphi(h_n) = c_1 h_n^{4k} + c_2\frac{1}{nh_n^{2m+1}}$ where $c_1 = c^2\|f^{(m+2k)}\|_{L_2}^2$ and $c_2 = \|M_k^{(m)}\|_{L_2}^2$ approximates IMSE. Minimization of φ yields the following optimal bandwidth:

$$h_{opt} = \left(\frac{(2m+1)c_2}{4knc_1}\right)^{1/(4k+2m+1)}. \tag{16}$$

3.3. Asymptotic Normality

In this section, we state a theorem that gives the asymptotic normality of our estimator under suitable normalization. The proof is omitted as it follows closely the proof of Theorem 9 in Mynbaev and Martins-Filho (2010).

Theorem 6. *Suppose that Assumption 3(a) and (b) hold. Let* $\int_{\mathbb{R}} |K^{(m)}(t)|^{2+\delta} dt < \infty$ *for some* $\delta > 0$. *If* $nh_n \to \infty$ *and* $nh_n^{4k+2m+1} \to 0$, *then*

$$(nh_n^{2m+1})^{1/2}(\hat{f}^{(m)}(x) - f^{(m)}(x)) \xrightarrow{d} N\left(0, f(x)\int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 dt\right). \tag{17}$$

4. MONTE CARLO STUDY

We implement our estimator $\hat{f}_k^{(m)}$ with $k = 1, 2, 3, 4, 8$ for derivatives of order $m = 0, 1, 2$. Our simulations were conducted for four different densities and three different seed kernels (K) to construct M_k. Since the fit quickly worsens as the derivative order grows, we did not consider $m > 2$. Note that if $k = 1$ and the seed K is symmetric, the kernel M_k is just K, so results reported below for $k = 1$ are actually for Rosenblatt–Parzen estimators. Mynbaev and Martins-Filho (2010) demonstrated that increasing k indeed improves estimation (they allowed k to take values 2, 4, 8) of the density. In case of derivative estimation, considering large k is technically more complex because the formula for the optimal bandwidth Eq. (16) requires a derivative of order $2k + m$ of the density.

The four densities to be estimated were proposed in Marron and Wand (1992) and are examples of normal mixtures. They are: (1) Gaussian ($f_1(x) \equiv N(1, 1)$); (2) Bimodal ($f_2(x) \equiv \frac{1}{2}N(-1, \frac{1}{9}) + \frac{1}{2}N(1, \frac{1}{9})$); (3) Separated-Bimodal ($f_3(x) \equiv \frac{1}{2}N(-1.5, \frac{1}{4}) + \frac{1}{2}N(1.5, \frac{1}{4})$), and (4) Trimodal ($f_4(x) \equiv \frac{9}{20}N(-\frac{6}{5}, \frac{9}{25}) + \frac{9}{20}N(\frac{6}{5}, \frac{9}{25}) + \frac{1}{10}N(0, \frac{1}{16})$). These four densities were used by Mynbaev and Martins-Filho (2010) for their simulations. They also considered one more density, whose second derivative is not continuous for all x but satisfies a Lipschitz condition of order 2. We exclude this fifth density from consideration because the optimal bandwidth we apply is not defined for the fifth density unless $m = 0$ and $k = 1$ (this case has already been considered by Mynbaev & Martins-Filho, 2010).

The three seeds we consider are: (1) Gaussian, (2) t distribution with 5 degrees of freedom, and (3) concentrated density. The concentrated density is defined as $\exp(-x^8)/c$, where $c = 2\Gamma(\frac{9}{8})$ is the normalization constant. Fig. 1 provides a graph for the different seeds we use. The motivation for the name of the concentrated density is clear from Fig. 1.

Note that the concentrated density has a flat top and nonexistent tails, while the t distribution is sharper at the top than the Gaussian and has fat tails. We did not use degrees of freedom larger than 5 for the t-distribution to avoid smoothing of the density at the top. The reason to experiment with different seeds was motivated by the fact that even with a very large number of observations ($n = 100,000$) the

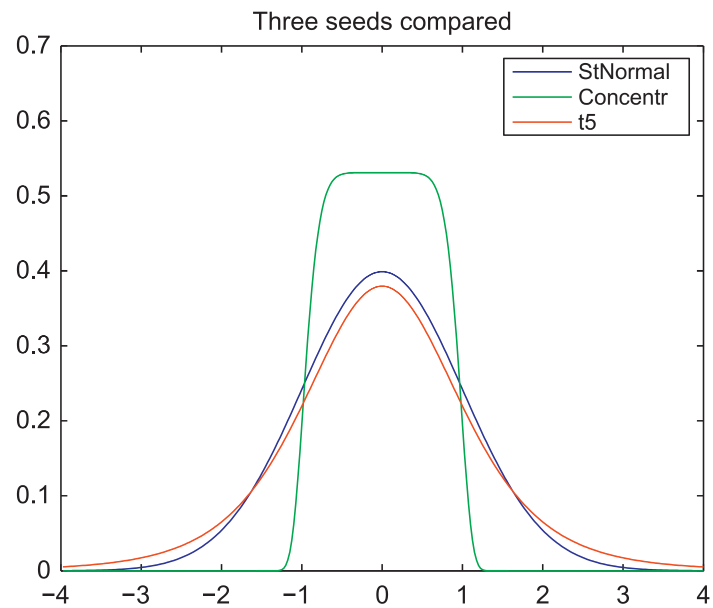

Fig. 1. Comparison of Three Seed Kernels.

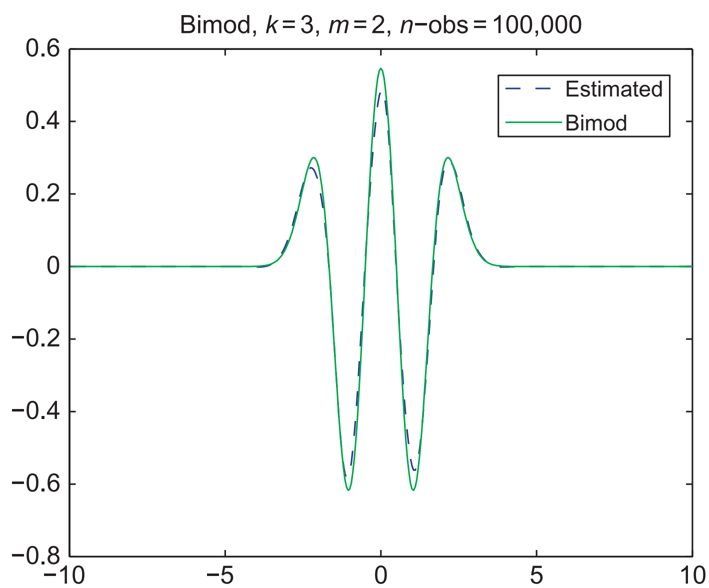

Fig. 2. Estimation of Second Derivative of Bimodal.

differences between the estimated density derivative and the true density derivative at the peaks and troughs of the graph do not vanish (see Fig. 2).

We report first the results for the Gaussian seed and then indicate the variations caused by replacing the seed. For each of the four densities 1,000 samples of size

Table 1. Five Estimators with Optimal Bandwidth h_{opt}.

	Estimators	$f_1(x)$		$f_2(x)$		$f_3(x)$		$f_4(x)$	
		B	MSE	B	MSE	B	MSE	B	MSE
	$n=200$								
$m=0$	$\hat{f}_1$	0.110	0.003	0.121	0.005	0.147	0.007	0.148	0.006
	$\hat{f}_2$	0.107	0.002	0.129	0.004	0.140	0.005	0.158	0.008
	$\hat{f}_3$	0.099	0.002	0.129	0.004	0.142	0.006	0.164	0.008
	$\hat{f}_4$	0.102	0.002	0.125	0.005	0.130	0.006	0.176	0.009
	$\hat{f}_8$	0.099	0.002	0.131	0.005	0.145	0.006	0.184	0.010
$m=1$	$\hat{f}_1$	0.202	0.009	0.338	0.030	0.463	0.060	0.571	0.088
	$\hat{f}_2$	0.177	0.008	0.293	0.026	0.428	0.055	0.666	0.123
	$\hat{f}_3$	0.170	0.009	0.296	0.024	0.455	0.051	0.762	0.170
	$\hat{f}_4$	0.176	0.007	0.301	0.022	0.443	0.052	0.841	0.203
	$\hat{f}_8$	0.178	0.008	0.299	0.025	0.440	0.056	0.931	0.267
$m=2$	$\hat{f}_1$	0.415	0.047	0.978	0.273	1.887	0.999	3.209	3.036
	$\hat{f}_2$	0.041	0.037	0.930	0.258	1.775	0.844	4.185	5.345
	$\hat{f}_3$	0.358	0.037	0.945	0.230	1.792	0.922	5.263	7.547
	$\hat{f}_4$	0.407	0.043	0.928	0.228	1.950	0.980	6.045	10.842
	$\hat{f}_8$	0.413	0.041	1.007	0.273	2.010	1.144	7.372	15.053
	$n=400$								
$m=0$	$\hat{f}_1$	0.086	0.002	0.101	0.003	0.114	0.004	0.111	0.004
	$\hat{f}_2$	0.075	0.002	0.095	0.003	0.107	0.003	0.116	0.004
	$\hat{f}_3$	0.077	0.001	0.093	0.003	0.107	0.003	0.123	0.004
	$\hat{f}_4$	0.077	0.001	0.094	0.002	0.106	0.003	0.128	0.005
	$\hat{f}_8$	0.075	0.001	0.097	0.003	0.106	0.003	0.130	0.005
$m=1$	$\hat{f}_1$	0.159	0.007	0.267	0.020	0.385	0.043	0.459	0.065
	$\hat{f}_2$	0.142	0.005	0.242	0.016	0.334	0.031	0.538	0.085
	$\hat{f}_3$	0.138	0.005	0.246	0.016	0.343	0.032	0.585	0.099
	$\hat{f}_4$	0.136	0.005	0.240	0.015	0.346	0.033	0.629	0.115
	$\hat{f}_8$	0.132	0.006	0.238	0.015	0.350	0.032	0.681	0.135
$m=2$	$\hat{f}_1$	0.343	0.032	0.848	0.194	1.621	0.741	2.840	2.421
	$\hat{f}_2$	0.317	0.027	0.880	0.164	1.514	0.640	3.353	3.252
	$\hat{f}_3$	0.305	0.024	0.773	0.160	1.480	0.604	4.068	4.688
	$\hat{f}_4$	0.326	0.027	0.775	0.151	1.570	0.679	4.924	7.040
	$\hat{f}_8$	0.322	0.025	0.791	0.165	1.534	0.651	5.472	8.532
	$n=600$								
$m=0$	$\hat{f}_1$	0.072	0.002	0.086	0.002	0.098	0.003	0.103	0.003
	$\hat{f}_2$	0.069	0.001	0.083	0.002	0.091	0.002	0.102	0.003
	$\hat{f}_3$	0.064	0.001	0.081	0.002	0.089	0.002	0.101	0.003
	$\hat{f}_4$	0.065	0.001	0.080	0.002	0.091	0.002	0.106	0.003
	$\hat{f}_8$	0.061	0.001	0.079	0.002	0.090	0.002	0.109	0.004
$m=1$	$\hat{f}_1$	0.148	0.006	0.240	0.016	0.348	0.036	0.414	0.050
	$\hat{f}_2$	0.129	0.004	0.210	0.013	0.302	0.025	0.457	0.060
	$\hat{f}_3$	0.124	0.004	0.209	0.012	0.309	0.026	0.501	0.072
	$\hat{f}_4$	0.130	0.004	0.206	0.011	0.305	0.024	0.547	0.084
	$\hat{f}_8$	0.116	0.003	0.205	0.010	0.312	0.027	0.574	0.096
$m=2$	$\hat{f}_1$	0.326	0.030	0.802	0.183	1.471	0.621	2.506	1.976
	$\hat{f}_2$	0.293	0.022	0.695	0.132	1.350	0.508	2.986	2.579
	$\hat{f}_3$	0.271	0.018	0.684	0.126	1.359	0.497	3.475	3.624
	$\hat{f}_4$	0.281	0.019	0.677	0.120	1.329	0.474	4.029	4.733
	$\hat{f}_8$	0.279	0.020	0.687	0.129	1.365	0.519	4.715	6.450

400 were generated. In our first set of simulations, five estimators were obtained for each sample: $\hat{f}_k(x)$ for $k = 1, 2, 3, 4, 8$ where, as stated earlier, $\hat{f}_1(x) = \hat{f}_{RP}(x)$, the Rosenblatt–Parzen estimator. For all estimators, the optimal bandwidth h_{opt} in Eq. (16) was used in our implementation. The usual caveat applies: in practice, this bandwidth is infeasible given that $f(x)$ is unknown. However, in the context of a Monte Carlo study it is desirable since estimation performance is not impacted by the noise introduced through a data-driven bandwidth selection. Table 1 provides average absolute bias (B) and average mean-squared error (MSE) for each density considered for $n = 200, 400, 600$ respectively. We observe the following general regularities. As follows from the theory, increases in the values of k seem to reduce average absolute bias and MSE, but this is not verified for all experiments. Specifically, the step from $k = 3$ to $k = 4$ does not always improve B and MSE in case of higher order derivatives or in case of densities that are more difficult to estimate, i.e, f_3 and f_4. Further, density functions with larger curvature (in increasing order of curvature f_1, f_2, f_3, and f_4) are more difficult to estimate both in terms of bias and MSE for all estimators considered. Our proposed estimators ($\hat{f}_2$, $\hat{f}_3$, $\hat{f}_4$) outperform the Rosenblatt–Parzen estimator both in terms of bias and MSE, except when estimating f_4.

For the other two seeds (concentrated and t distribution), we give a verbal description of the simulation results (full tables are available on request). When we use the concentrated distribution as a seed, the statistics behave the same as one moves right along the table (they worsen when the density curvature increases). The behavior along columns changes. For $m = 0$, the case $k = 1$ is about as good as $k = 2$, except for f_4 when the Rosenblatt–Parzen is the best. In case $m = 1, 2$, the Rosenblatt–Parzen outperforms the others. This is true both for B and MSE. With t distribution with 5 degrees of freedom as a seed, the requirement $\int K(t)t^{2k}dt < \infty$ implied by the definition of the optimal bandwidth limits the value of k: $k \le 2$. For $m = 1, 2$ our estimator with $k = 2$ outperforms the Rosenblatt–Parzen for all densities, except f_4. For $m = 0$, our estimator is better everywhere.

Finally, if we compare, for fixed k and m, the three seeds, the Gaussian is the best of all, with the margins being the largest for f_4, which is the most difficult to estimate. The Gaussian density seems to strike the right balance between concentration and dispersion.

5. CONCLUSION

We have shown that taking derivatives of order m of the density estimators in the class first proposed by Mynbaev and Martins-Filho (2010) produce estimators for the m-order derivative of the densities that have desirable asymptotic properties.

In particular, these estimators are (uniformly) consistent and asymptotically normally distributed under suitable normalization. In addition, the reduction in the order of the bias, relative to the classical Rosenblatt–Parzen density estimator, first discovered in Mynbaev and Martins-Filho (2010) in the context of density estimation, also manifests itself in the context of derivative estimation. These theoretical results are supported by a small Monte Carlo study, but in agreement with previous simulations we conducted in the case of density estimation, very large values of k seem, in some contexts, to damage finite sample performance as measured by MSE. An interesting extension of this research would be to develop a practical criterion for the selection of k, viz., a criterion for the selection of an optimal density or density derivative estimator in the class we have proposed. We leave such efforts for future studies.

NOTES

1. For an example of one such density, see Mynbaev and Martins-Filho (2010, p. 232).
2. Theorem 1 in Mynbaev and Martins-Filho (2010) obtained expressions for $H(x)$ and $\varepsilon(x)$ for the Gaussian and Cauchy densities.
3. This proof extends the arguments from Besov, Il'in, and Nikolskii (1975, p. 254).
4. Note that since x_0 and $2k$ are fixed, the operator acts on the function g and the result is a function of x.

ACKNOWLEDGMENTS

We thank two anonymous referees for useful comments.

REFERENCES

Besov, O., Il'in, V., & Nikolskii, S. (1975). *Integral representations of functions and imbedding theorems*. VSP BV, Nauka: Moscow.

Bhattacharya, P. K. (1967). Estimation of a probability density function and its derivatives. *Sankhya, 29*, 373–382.

Härdle, W., & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association 20, 986–995.

Henderson, D., & Parmeter, C. (2012). Canonical higher-order kernels for density derivative estimation. *Statistics and Probability Letters, 82*, 1383–1387.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics, 40*(2), 633–643.

Marron, J. S., & Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics, 20*, 712–736.

Muller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and models. *Annals of Statistics, 12*, 766–774.

Mynbaev, K., & Martins-Filho, C. (2010). Bias reduction in kernel density estimation via Lipschitz condition. *Journal of Nonparametric Statistics, 22*, 219–235.

Schuster, E. F. (1969). Estimation of a probability density and its derivatives. *Annals of Mathematical Statistics, 40*, 1187–1196.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics, 6*, 177–184.

Singh, R. (1977). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Annals of Statistics, 5*, 394–399.

Singh, R. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika, 66*, 177–180.

Singh, R. (1987). Mise of kernel estimates of a density and its derivatives. *Statistics and Probability Letters, 5*, 153–159.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. New York, NY: Springer-Verlag.

APPENDIX A: LEMMAS AND PROOFS

Lemma 1. *(a) Let* $\gamma_i = \sum_{s=-k}^{k} c_{k,s} s^i$ *for* $i = 0, \ldots, 2k$. *Then,* $\gamma_0 = \cdots = \gamma_{2k-1} = 0$, $\gamma_{2k} = (-1)^{2k}(2k)!$; *(b) Suppose* K *has finite moments* $\alpha_l(K)$ *of orders* $l \le 2k$. *Then* M_k *has moments* $\alpha_l(M_k) = 0 < l < 2k$, $\alpha_{2k}(M_k) = -\frac{\gamma_{2k}}{c_{k,0}}\,\alpha_{2k}(K)$.

Proof. (a) The function $\phi_q(x) = (1-x)^q$ vanishes at $x = 1$ together with all its derivatives of orders $l < q$. For the qth derivative, we have $\phi_q^{(q)}(x) = \left(\frac{d}{dx}\right)^q [\phi_q(x)] = (-1)^q q!$. Now, consider the linear operator $\left(x\frac{d}{dx}\right)$. By induction, for $l < q$ we have $\left(x\frac{d}{dx}\right)^l [\phi_q(x)] = \sum_{j=1}^{l} a_{j,l} x^j \phi_q^{(j)}(x)$, where $a_{j,l}$ are constants and $a_{l,l} = 1$. Now it is easy to see that

$$\left(x\frac{d}{dx}\right)^l [\phi_q(x)]|_{x=1} = \begin{cases} 0 & \text{if } l < q \\ (-1)^q q! & \text{if } l = q \end{cases}. \tag{A.1}$$

Note that by the binomial theorem, $\phi_q(x) = \sum_{m=0}^{q} (-1)^m x^m C_q^m$ and we see that

$$\left(x\frac{d}{dx}\right)^l [\phi_q(x)] = \sum_{m=0}^{q} (-1)^m m^l x^m C_q^m. \tag{A.2}$$

Comparing (A.1) and (A.2) we have

$$\sum_{m=0}^{q} (-1)^m m^l C_q^m = \begin{cases} 0 & \text{if } l < q \\ (-1)^q q! & \text{if } l = q \end{cases}. \tag{A.3}$$

By the definition of $c_{k,s}$ and replacing $s + k$ with m, we have

$$\begin{aligned}\gamma_i &= \sum_{s=-k}^{k} (-1)^{s+k} C_{2k}^{s+k} s^i = \sum_{m=0}^{2k} (-1)^m C_{2k}^m (m-k)^i \\ &= \sum_{m=0}^{2k} (-1)^m C_{2k}^m \sum_{j=0}^{i} C_i^j m^j (-k)^{i-j} = \sum_{j=0}^{i} C_i^j (-k)^{i-j} \sum_{m=0}^{2k} (-1)^m C_{2k}^m m^j.\end{aligned}$$

This identity and (A.3) prove statement (a).[3] b) Replacing $x/s = t$ we have

$$\begin{aligned}\int_R M_k(x)x^l dx &= -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}s^l\int_R K\left(\frac{x}{s}\right)\left(\frac{x}{s}\right)^l dx\\ &= -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}s^l c_{k,s}\int_R K(t)\,t^l dx\\ &= -\frac{1}{c_{k,0}}\sum_{s=-k}^{k}s^l c_{k,s}\alpha_l(K) = -\frac{1}{c_{k,0}}\gamma_l\alpha_l(K)\end{aligned}$$

which completes the proof. ■

Lemma 2. *Let $I(x,g) = \frac{1}{(2k)!}\int_{x_0}^{x}(x-t)^{2k}g(t)dt$ denote the $(2k+1)$-fold integration operator for some fixed x_0 and Assumption 1 hold.*[4] *If the density f has an absolutely continuous derivative $f^{(m+2k)}$, then the bias of $\hat{f}^{(m)}(x)$ has the representation*

$$B(\hat{f}^{(m)}(x)) = -\frac{\gamma_{2k}}{c_{k,0}}\frac{f^{(m+2k)}(x_0)}{(2k)!}h^{2k}\alpha_{2k}(K) - \frac{1}{c_{k,0}}\int_{\mathbb{R}}K(t)\Delta_{ht}^{2k}I(x, f^{(m+2k+1)})dt \tag{A.4}$$

where $\alpha_{2k}(K) = \int_{\mathbb{R}}K(t)t^{2k}dt$ and x_0 is arbitrary.

Proof. If $g^{(2k)}$ is absolutely continuous, then $g^{(2k+1)}$ is summable and by Taylor's theorem with remainder in integral form one has $g(x) = \sum_{i=0}^{2k}\frac{g^{(i)}(x_0)}{i!}(x-x_0)^i + \frac{1}{(2k)!}\int_{x_0}^{x}(x-t)^{2k}g^{(2k+1)}(t)dt$.

Applying this formula to $g = f^{(m)}$ and recalling our notation for the integration operator, we get

$$f^{(m)}(x) = \sum_{i=0}^{2k}\frac{f^{(m+i)}(x_0)}{i!}(x-x_0)^i + I(x, f^{(m+2k+1)}).$$

In view of Eq. (A.9), we need to consider

$$\Delta_{ht}^{2k}f^{(m)}(x) = \sum_{i=0}^{2k}\frac{f^{(m+i)}(x_0)}{i!}\Delta_{ht}^{2k}(x-x_0)^i + \Delta_{ht}^{2k}I(x, f^{(m+2k+1)}) \tag{A.5}$$

where the difference is applied with respect to the variable x, x_0 being fixed. By Lemma 1

$$\Delta_{ht}^{2k}(x-x_0)^i = \sum_{s=-k}^{k} c_{k,s}(x+sht-x_0)^i = \sum_{s=-k}^{k} c_{k,s} \sum_{j=0}^{i} C_i^j (x-x_0)^{i-j}(sht)^j$$
$$= \sum_{j=0}^{i} C_i^j (x-x_0)^{i-j}(ht)^j \sum_{s=-k}^{k} c_{k,s}s^j = \begin{cases} 0, & i<2k \\ \gamma_{2k}(ht)^{2k}, & i=2k \end{cases} \quad \text{(A.6)}$$

Under Assumption 1, combining Eqs. (A.9), (A.5), and (A.6) we finish the proof of Eq. (A.4):

$$B(\hat{f}^{(m)}(x)) = -\frac{1}{c_{k,0}}\int_{\mathbb{R}} K(t)\left[\frac{f^{(m+2k)}(x_0)}{(2k)!}\gamma_{2k}(ht)^{2k} + \Delta_{ht}^{2k} I(x, f^{(m+2k+1)})\right]dt$$
$$= -\frac{\gamma_{2k}}{c_{k,0}(2k)!} f^{(m+2k)}(x_0)h^{2k}\int_{\mathbb{R}} K(t)t^{2k}dt$$
$$- \frac{1}{c_{k,0}}\int_{\mathbb{R}} K(t)\Delta_{ht}^{2k} I(x, f^{(m+2k+1)})dt.$$

■

Theorem 1. Under Assumption 1, we show that $\max\{|K(s)|, |K^{(1)}(s)|, \ldots, |K^{(m-1)}(s)|\} = o\left(\frac{1}{|s|}\right)$ as $|s| \to \infty$. Let $s>0$. It is well-known that the Sobolev space $W_1^m[0,1]$ is embedded in $C^j[0,1]$ for $j=0,1,\ldots,m-1$, that is, with some constant c independent of K one has $\|K^{(j)}\|_{C[0,1]} \le c\int_0^1 \left(|K(t)| + |K^{(m)}(t)|\right)dt$. Applying this bound to the segment $[s, s+1]$ and using the fact that $|t/s| \ge 1$ for $t \in [s, s+1]$ we get

$$\max_j |K^{(j)}(s)| \le c\int_s^{s+1} \left(|K(t)| + |K^{(m)}(t)|\right)dt$$
$$\le \frac{c}{|s|}\int_s^{s+1} \left(|K(t)| + |K^{(m)}(t)|\right)|t|dt.$$

The case of $s<0$ is treated similarly.

Under Assumption 1 $\left|K^{(j)}(t)f^{(m-1-j)}(x-sh_nt)\right| = o\left(\frac{1}{|t|}\right)O(|x-sh_nt|) = o\left(\left|\frac{x-sh_nt}{t}\right|\right) = o(1)$, as $|t|\to\infty$ for $j=0,\ldots,m-1$, $h_n>0$. Therefore, we can integrate by parts, and from Eq. (5) and a change of variables, we obtain

$$
\begin{aligned}
E(\hat{f}_k^{(m)}(x)) = E(u_1) &= \frac{1}{h_n^{m+1}} \int_{-\infty}^{+\infty} M_k^{(m)}\left(\frac{x-t}{h_n}\right) f(t)dt \\
&= \frac{1}{h_n^m} \int_{-\infty}^{+\infty} M_k^{(m)}(l) f(x-h_n l)dl \\
&= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m h_n^m} \int_{-\infty}^{+\infty} K^{(m)}\left(\frac{l}{s}\right) f(x-h_n l)dl \\
&= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m} \left[\frac{s}{h_n^m} K^{(m-1)}\left(\frac{l}{s}\right) f(x-h_n l)\Big|_{-\infty}^{+\infty} \right. \\
&\quad \left. + \frac{s}{h_n^{m-1}} \int_{-\infty}^{+\infty} K^{(m-1)}\left(\frac{l}{s}\right) f^{(1)}(x-h_n l)dl \right] \\
&= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m} \left[\frac{s^2}{h_n^{m-1}} K^{(m-2)}\left(\frac{l}{s}\right) f'(x-h_n l)\Big|_{-\infty}^{+\infty} \right. \\
&\quad \left. + \frac{s^2}{h_n^{m-2}} \int_{-\infty}^{+\infty} K^{(m-2)}\left(\frac{l}{s}\right) f^{(1)}(x-h_n l)dl \right] \\
&= \cdots = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m} \left[\frac{s^m}{h_n} K\left(\frac{l}{s}\right) f^{(m-1)}(x-h_n l)\Big|_{-\infty}^{+\infty} \right. \\
&\quad \left. + \frac{s^m}{h_n} h_n \int_{-\infty}^{+\infty} K\left(\frac{l}{s}\right) f^{(m)}(x-h_n l)dl \right] \\
&= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} \int_{-\infty}^{+\infty} K\left(\frac{l}{s}\right) f^{(m)}(x-h_n l)dl \\
&= -\frac{1}{c_{k,0}} \left[\sum_{s=-k}^{-1} \frac{c_{k,s}}{-s}(-s) \int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt \right. \\
&\quad \left. + \sum_{s=1}^{k} \frac{c_{k,s}}{s} s \int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt \right] \\
&= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt. \qquad \text{(A.7)}
\end{aligned}
$$

Hence, from Eqs. (1), (2), and (A.7) we obtain

$$
\begin{aligned}
B(\hat{f}^{(m)}(x)) &= -\frac{1}{c_{k,0}}\left[\sum_{s=-k}^{-1} c_{k,s}\int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt\right. \\
&\qquad \left. + \sum_{s=1}^{k} c_{k,s}\int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt\right] - \frac{c_{k,0}}{c_{k,0}} f^{(m)}(x) \qquad \text{(A.8)} \\
&= -\frac{1}{c_{k,0}}\left[\sum_{|s|=1}^{k} c_{k,s}\int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+sh_n t)dt\right. \\
&\qquad \left. + c_{k,0}\int_{-\infty}^{+\infty} K(-t) f^{(m)}(x+0h_n t)dt\right] \\
&= -\frac{1}{c_{k,0}}\int_{-\infty}^{+\infty} K(-t)\Delta_{h_n t}^{2k} f^{(m)}(x)dt \\
&= -\frac{1}{c_{k,0}}\int_{-\infty}^{+\infty} K(t)\Delta_{h_n t}^{2k} f^{(m)}(x)dt, \qquad \text{(A.9)}
\end{aligned}
$$

where the last equality follows from the symmetry of K. ■

Theorem 2. (a) We denote $\psi_j = \frac{x-X_j}{h_n}$, then we can rewrite Eq. (4) as $\hat{f}_k^{(m)}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^{m+1}} M_k^{(m)}(\psi_j)$ and using Eq. (5), we get

$$
M_k^{(m)}(\psi_j) = -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|s^m} K^{(m)}\left(\frac{\psi_j}{s}\right). \qquad \text{(A.10)}
$$

Under Assumption 2(a) the inversion theorem for Fourier transforms gives

$$
K^{(m)}\left(\frac{\psi_j}{s}\right) = \frac{(-i)^{(m)}}{2\pi}\int_{\mathbb{R}} \exp\left\{\frac{-it\psi_j}{s}\right\} t^m \phi_K(t)dt. \qquad \text{(A.11)}
$$

Using Eqs. (4), (5), (A.10), and (A.11) and by changing variables of integration, we have

$$
\begin{aligned}
\hat{f}_k^{(m)}(x) &= \frac{1}{nh_n^{m+1}} \sum_{j=1}^{n} M_k^{(m)}\left(\frac{x - X_j}{h_n}\right) = \frac{1}{nh_n^{m+1}} \sum_{j=1}^{n} M_k^{(m)}(\psi_j) \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{nh_n^{m+1}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m} \int_{\mathbb{R}} \exp\left\{\frac{-it\psi_j}{s}\right\} t^m \phi_K(t)dt \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{nh_n^{m+1}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s^m} \int_{\mathbb{R}} \exp\left\{-it\left(\frac{x - X_j}{sh}\right)\right\} t^m \phi_K(t)dt \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{n} \sum_{|s|=1}^{k} c_{k,s} \int_{\mathbb{R}} \exp\{-i\gamma x\} \exp\{i\gamma X_j\} \gamma^m \phi_K(sh_n\gamma)d\gamma \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \int_{\mathbb{R}} \exp\{-i\gamma x\} \sum_{j=1}^{n} \frac{1}{n} \exp\{i\gamma X_j\} \sum_{|s|=1}^{k} c_{k,s}\gamma^m \phi_K(sh_n\gamma)d\gamma \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \int_{\mathbb{R}} \exp\{-i\gamma x\}\hat{\phi}_n(\gamma)\Delta(\gamma)d\gamma
\end{aligned}
$$

where $\hat{\phi}_n(\gamma) = \sum_{j=1}^{n} \frac{1}{n} \exp\{i\gamma X_j\}$ is an unbiased estimator for the characteristic function $\phi_f(t)$ of f and $\Delta(\gamma) = \sum_{|s|=1}^{k} c_{k,s}\gamma^m \phi_K(sh_n\gamma)$. Thus,

$$
\begin{aligned}
E(\hat{f}_k^{(m)}(x)) &= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \int_{\mathbb{R}} \exp\{-i\gamma x\} E\hat{\phi}_n(\gamma)\Delta(\gamma)d\gamma \\
&= -\frac{(-i)^{(m)}}{2\pi c_{k,0}} \int_{\mathbb{R}} \exp\{-i\gamma x\}\phi_f(\gamma)\Delta(\gamma)d\gamma
\end{aligned}
$$

so that $|\hat{f}_k^{(m)}(x) - E(\hat{f}_k^{(m)}(x))| \le c \int_{\mathbb{R}} |\hat{\phi}_n(\gamma) - \phi_f(\gamma)||\exp\{-i\gamma x\}||\Delta(\gamma)|d\gamma$. But since $|\exp\{-i\gamma x\}| = 1$,

$$
\sup_{x\in\mathbb{R}} |\hat{f}_k^{(m)}(x) - E(\hat{f}_k^{(m)}(x))| \le c \int_{\mathbb{R}} |\hat{\phi}_n(\gamma) - \phi_f(\gamma)||\Delta(\gamma)|d\gamma.
$$

with no sup on the right-hand side because it does not depend on x. It follows from Lemma 2.1 of Jennrich (1969) that $\sup|\cdot|$ is measurable, its expectation is well defined and

$$
E\left(\sup_{x\in\mathbb{R}} |\hat{f}_k^{(m)}(x) - E\hat{f}_k^{(m)}(x)|\right) \le c \int_{\mathbb{R}} E\left|\hat{\phi}_n(\gamma) - \phi_f(\gamma)\right| |\Delta(\gamma)|d\gamma.
$$

Now,

$$E\left(|\hat{\phi}_n(\gamma)-\phi_f(\gamma)|\right)=E\left(\left|\frac{1}{n}\sum_{j=1}^{n}\exp\{i\gamma X_j\}-E(\exp\{i\gamma X_j\})\right|\right)$$

$$=E(|Y_1+iY_2|)=E|Y_1|+E|Y_2|=E[(Y_1^2+Y_2^2)]^{\frac{1}{2}}$$

$$\le[E(Y_1^2+Y_2^2)]^{\frac{1}{2}}\le(EY_1^2)^{\frac{1}{2}}+(EY_2^2)^{\frac{1}{2}}$$

where

$$\begin{cases} Y_1=\frac{1}{n}\sum\limits_{j=1}^{n}(\cos(\gamma X_j)-E(\cos(\gamma X_j))) \\ Y_2=\frac{1}{n}\sum\limits_{j=1}^{n}(\sin(\gamma X_j)-E(\sin(\gamma X_j))). \end{cases}$$

Using the i.i.d. assumption, it is easy to see that

$$EY_1^2=\frac{1}{n^2}\sum_{j=1}^{n}\left[E\cos^2(\gamma X_j)-(E\cos(\gamma X_j))^2\right]$$

$$=\frac{1}{n^2}\sum_{j=1}^{n}V\left(\cos(\gamma X_j)\right)=\frac{1}{n}\left[V(\cos(\gamma X_1)\right]$$

$$EY_2^2=\frac{1}{n^2}\sum_{j=1}^{n}\left[E\sin^2(\gamma X_j)-(E\sin(\gamma X_j))^2\right]$$

$$=\frac{1}{n^2}\sum_{j=1}^{n}V\left(\sin(\gamma X_j)\right)=\frac{1}{n}\left[V(\sin(\gamma X_1))\right]$$

Consequently, $(EY_1^2)^{\frac{1}{2}}+(EY_2^2)^{\frac{1}{2}}=\left(\frac{1}{n}V(\cos(\gamma X_1)\right)^{\frac{1}{2}}+\left(\frac{1}{n}V(\sin(\gamma X_1)\right)^{\frac{1}{2}}$. Since $E\cos^2(\gamma X_1)\le 1$ and $E\sin^2(\gamma X_1)\le 1$, $V(\cos(\gamma X_1))^{\frac{1}{2}}=\left[E\cos^2(\gamma X_1)\right.$ $\left.-(E\cos(\gamma X_1))^2\right]^{\frac{1}{2}}\le\left[E\cos^2(\gamma X_1)+E\cos^2(\gamma X_1)\right]^{\frac{1}{2}}\le\sqrt{2}$. $V(\sin(\gamma X_1))^{\frac{1}{2}}$ $=\left[E\sin^2(\gamma X_1)-(E\sin(\gamma X_1))^2\right]^{\frac{1}{2}}\le\left[E\sin^2(\gamma X_1)+E\sin^2(\gamma X_1)\right]^{\frac{1}{2}}$ $\le\sqrt{2}$. $(EY_1^2)^{\frac{1}{2}}+(EY_2^2)^{\frac{1}{2}}=\frac{2\sqrt{2}}{\sqrt{n}}$. Hence, $E\left(\left|\hat{\phi}_n(\gamma)-\phi_f(\gamma)\right|\right)\le\frac{2\sqrt{2}}{\sqrt{n}}$ and

$$\int_{\mathbb{R}}|\Delta(\gamma)|d\gamma\le\sum_{|s|=1}^{k}|c_{k,s}|\int_{\mathbb{R}}|\gamma^m||\phi_K(sh_n\gamma)|d\gamma$$

$$\le\frac{1}{h_n^{m+1}}\sum_{|s|=1}^{k}\frac{|c_{k,s}|}{s^{m+1}}\int_{\mathbb{R}}|t^m\phi_K(t)|dt=\frac{c}{h_n^{m+1}}\int_{\mathbb{R}}|t^m\phi_K(t)|dt.$$

Finally, $E\left(\sup_{x\in\mathbb{R}}|\hat{f}_k^{(m)}(x)-E\hat{f}_k^{(m)}(x)|\right)\le\frac{c}{h_n^{m+1}\sqrt{n}}\int_{\mathbb{R}}|t^m\phi_K(t)|dt$ which tends to zero as $n\to\infty$ under Assumption 2(c) ($nh_n^{2m+2}\to\infty$). Further, by Markov's inequality

$$P\left(\sup_x|\hat{f}_k^{(m)}(x)-E(\hat{f}_k^{(m)}(x))|>\varepsilon\right)\to 0 \tag{A.12}$$

as $n\to\infty$ for all $\varepsilon>0$. Therefore, $\sup_{x\in\mathbb{R}}|\hat{f}_k^{(m)}(x)-E\hat{f}_k^{(m)}(x)|\xrightarrow{p}0$. Note that

$$\sup_{x\in\mathbb{R}}|\hat{f}_k^{(m)}(x)-f^{(m)}(x)|\le\sup_{x\in\mathbb{R}}|\hat{f}_k^{(m)}(x)-E\hat{f}_k^{(m)}(x)|+\sup_{x\in\mathbb{R}}|E\hat{f}_k^{(m)}(x)-f^{(m)}(x)|.$$

The first term on the right-hand side of the inequality is uniformly $o_p(1)$ from Eq. (A.12). The second term tends to zero by Eq. (A.7), Assumption 2(b) and Theorem 5 (for the case where $m=0$) in Mynbaev and Martins-Filho (2010). We have $\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|\hat{f}_k^{(m)}(x)-f^{(m)}(x)|=0$. Consequently, $\hat{f}_k^{(m)}(x)$ is uniformly consistent. ■

Theorem 3. Assumption 3(c) implies for any $N>0$

$$\int_{|t|>N}|K(t)|\,dt\le\int_{|t|>N}|K(t)|\left|\frac{t}{N}\right|^{2k}dt\le N^{-2k}\int_{-\infty}^{\infty}|K(t)|\,t^{2k}dt. \tag{A.13}$$

Then, using Eq. (A.9) and Assumption 3(b), we have

$$\begin{aligned}\left|B(\hat{f}_k^{(m)}(x))\right|&=\left|\frac{1}{c_{k,0}}\right|\left|\int_{-\infty}^{\infty}K(t)\Delta_{h_nt}^{2k}f^{(m)}(x)dt\right|\\&\le c_1\left(\int_{|h_nt|\le\varepsilon_{2k}(x)}+\int_{|h_nt|>\varepsilon_{2k}(x)}\right)|K(t)||\Delta_{h_nt}^{2k}f^{(m)}(x)|dt\\&\le c_2\left[H_{2k,m}(x)\int_{|h_nt|\le\varepsilon_{2k}(x)}|K(t)|(h_nt)^{2k}dt\right.\\&\qquad\left.+\sup_{x\in\mathbb{R}}|f^{(m)}(x)|\int_{|h_nt|>\varepsilon_{2k}(x)}|K(t)|dt\right].\end{aligned}$$

It remains to apply Eqs. (A.13) and (8) to obtain Eq. (9).

Now, we proceed with derivation of Eq. (11). According to Eq. (7) we need to evaluate $E(u_1^2)$ and $E(u_1)^2$. By Eqs. (6) and (9),

$$Eu_1 = E(\hat{f}_k^{(m)}(x)) = f^{(m)}(x) + B(\hat{f}_k^{(m)}(x)) = f^{(m)}(x) + R_{2k}(x, h_n, m) \tag{A.14}$$

where $R_{2k}(x, h_n)$ satisfies Eq. (12). Now, $E(u_1^2) = \left(\frac{1}{h_n^{m+1}}\right)^2 \int_{\mathbb{R}} \left[M_k^{(m)}\left(\frac{x-t}{h_n}\right)\right]^2 f(t)dt = \frac{1}{h_n^{2m+1}} \int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 f(x - h_n t)dt$. Consider

$$\begin{aligned}
&\int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 f(x - h_n t)dt - f(x)\int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 dt \\
&\quad = \int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 [f(x - h_n t) - f(x)]\, dt
\end{aligned}$$

Then, similarly to Eq. (A.13), we have

$$\int_{|t|>N} \left|M_k^{(m)}(t)\right|^2 dt \le \int_{|t|>N} \left|M_k^{(m)}(t)\right|^2 \left|\frac{t}{N}\right| dt \le N^{-1} \int_{-\infty}^{\infty} \left|M_k^{(m)}(t)\right|^2 |t|dt. \tag{A.15}$$

Using Eqs. (10) and (A.15), we have

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} \left(M_k^{(m)}(t)\right)^2 (f(x - h_n t) - f(x))dt\right| \\
&\le \left(\int_{|h_n t| \le \varepsilon_1(x)} + \int_{|h_n t| > \varepsilon_1(x)}\right) \left|M_k^{(m)}(t)\right|^2 |f(x - h_n t) - f(x)|\, dt \\
&\le H_1(x) \int_{|h_n t| \le \varepsilon_1(x)} \left|M_k^{(m)}(t)\right|^2 |(h_n t)|dt + \sup_{x\in\mathbb{R}} |f(x)| \int_{|h_n t| > \varepsilon_1(x)} \left|M_k^{(m)}(t)\right|^2 dt.
\end{aligned}$$

Then using Eq. (10), we get

$$\int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 f(x - h_n t)dt = f(x) \int_{\mathbb{R}} \left[M_k^{(m)}(t)\right]^2 dt + R_1(x, h_n) \tag{A.16}$$

where $R_1(x, h_n)$ satisfies Eq. (12).

Now we show that $\int_{\mathbb{R}}\left[M_k^{(m)}(t)\right]^2 dt < \infty$. From Eq. (5), we have $M_k^{(m)}(x) = \sum_{|s|=1}^{k} a_s K^{(m)}\left(\frac{x}{s}\right)$, where $a_s = -\frac{1}{c_{k,0}}\frac{c_{k,s}}{|s|}s^{(-m)}$. Hence, by Hölder's inequality

$$\begin{aligned}\int_{\mathbb{R}}\left(M_k^{(m)}(x)\right)^2 dx &= \int_{\mathbb{R}}\sum_{|s|,|t|=1}^{k} a_s a_t K^{(m)}\left(\frac{x}{s}\right)K^{(m)}\left(\frac{x}{t}\right)dx\\ &\le \sum_{|s|,|t|=1}^{k}|a_s a_t|\int_{\mathbb{R}}\left|K^{(m)}\left(\frac{x}{s}\right)\right|\left|K^{(m)}\left(\frac{x}{t}\right)\right|dx\\ &\le \sum_{|s|,|t|=1}^{k}|a_s a_t|\left(\int_{\mathbb{R}}\left|K^{(m)}\left(\frac{x}{s}\right)\right|^2 dx\right)^{\frac{1}{2}}\left(\int_{\mathbb{R}}\left|K^{(m)}\left(\frac{x}{t}\right)\right|^2 dx\right)^{\frac{1}{2}}\\ &= c_1\left(\int_{\mathbb{R}}\left|K^{(m)}(t)\right|^2 dt\right) < \infty\end{aligned}$$

because $K^{(m)} \in L_2(\mathbb{R})$.

Note that Eq. (11) is a consequence of Eqs. (A.14) and (A.16). In addition, if $f(x) \neq 0$ and for small h_n we can rewrite Eq. (11) as

$$V(\hat{f}_k^{(m)}(x)) = \frac{1}{nh_n^{2m+1}}\left[f(x)\int_{\mathbb{R}}\left(M_k^{(m)}(t)\right)^2 dt + O(h_n)\right]. \tag{A.17}$$

■

Theorem 4. By Eq. (7) in Mynbaev and Martins-Filho (2010) $\left|\Delta_h^{2k}g(x)\right| \le h^{2k}\sup_{|x-t|\le k|h|}|g^{(2k)}(t)|$. It is easy to verify that $\left(\frac{d}{dx}\right)^{2k} I(x,g) = \int_{x_0}^{x} g(t)dt$. Hence, using the last equation and the preceding inequality, we have

$$\begin{aligned}\left|\Delta_{ht}^{2k} I(x, f^{(m+2k+1)})\right| &\le (ht)^{2k}\sup_{|x-y|\le k|ht|}\left|\left(\frac{d}{dy}\right)^{2k} I(y, f^{(m+2k+1)})\right|\\ &= (ht)^{2k}\sup_{|x-y|\le k|ht|}\left|\int_{x_0}^{y} f^{(m+2k+1)}(z)dz\right|.\end{aligned}$$

Next, given that x_0 is arbitrary, we set $x_0 = x$ and use Eq. (A.6) and Assumption 4(b) to obtain

$$\begin{aligned}\left|\Delta_{ht}^{2k} I(x, f^{(m+2k+1)})\right| &\le (ht)^{2k}\sup_{|x-y|\le k|ht|}\left|f^{(m+2k)}(y) - f^{(m+2k)}(x)\right|\\ &\le \begin{cases}|ht|^{2k+1}kH_{1,m}(x) & \text{if } k|ht| \le \varepsilon(x),\\ 2(ht)^{2k}\|f^{(m+2k)}\|_C & \text{if } k|ht| > \varepsilon(x).\end{cases}\end{aligned}$$

It follows that

$$\begin{aligned}\left|\int_{\mathbb{R}} K(t)\Delta_{ht}^{2k} I(x, f^{(m+2k+1)})dt\right| \le\ & kh_n^{2k+1} H_{1,m}(x) \int_{k|ht|\le\varepsilon(x)} |K(t)||t|^{2k+1}dt \\ &+ 2h_n^{2k}\|f^{(m+2k)}\|_C \\ &\times \int_{k|ht|>\varepsilon(x)} |K(t)|t^{2k}dt.\end{aligned}$$

In the first integral on the right expand the domain of integration; in the second one use the inequality $1 < k|ht|/\varepsilon(x)$ and then expand the domain. The outcome is

$$\begin{aligned}\left|\int_{\mathbb{R}} K(t)\Delta_{ht}^{2k} I(x, f^{(m+2k+1)})dt\right| \le\ & kh_n^{2k+1} H_{1,m}(x)\beta_{2k+1}(K) \\ &+ 2kh_n^{2k+1}\|f^{(m+2k)}\|_C \frac{1}{\varepsilon(x)}\beta_{2k+1}(K) \\ \le\ & kh_n^{2k+1}\beta_{2k+1}(K)\left[H_{1,m}(x) + \frac{2\|f^{(m+2k)}\|_C}{\varepsilon(x)}\right].\end{aligned}$$

This equation and Eq. (A.4) prove the theorem. ■

Theorem 5. Under the conditions imposed, Eq. (A.17) implies

$$\int_{\mathbb{R}} V(\hat{f}^{(m)}(x))dx = \frac{1}{nh_n^{2m+1}}\left[\int_{\mathbb{R}} f(x)dx\|M_k^{(m)}\|_{L_2}^2 + O(h_n)\right],$$

while Eq. (14) gives $\int_{\mathbb{R}}\left[B(\hat{f}^{(m)}(x))\right]^2 dx = h^{4k}\left[c^2\|f^{(m+2k)}\|_{L_2}^2 + O(h)\right]$. Summing the last two equations we get Eq. (15). The rest is obvious and

$$h_{opt} = \left\{\frac{2m+1}{4kn\gamma_{2k}^2}\frac{\int [M_k^{(m)}(t)]^2dt}{\int [f^{(m+2k)}]^2dt(\int K(t)t^{2k}dt)^2}\left(\frac{(2k)!}{k!}\right)^4\right\}^{\frac{1}{4k+2m+1}}. \qquad \text{(A.18)}$$

■