STAPRO: 7780 Model 3Gsc pp. 1–6 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Q1 Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view[☆]

Q2 Kairat Mynbaev[a], Carlos Martins-Filho[b,c,*]

[a] International School of Economics, Kazakh-British Technical University, Tolebi 59, Almaty 050000, Kazakhstan

[b] Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA

[c] IFPRI, 2033 K Street NW, Washington, DC 20006-1002, USA

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 19 November 2016
Accepted 20 November 2016
Available online xxxx

MSC:
62G07
62G10
62G20

Keywords:
Kernel density estimation
Higher order kernels
Bias reduction

ABSTRACT

We define a new bandwidth-dependent kernel density estimator that improves existing convergence rates for the bias, and preserves that of the variation, when the error is measured in L_1. No additional assumptions are imposed to the extant literature.

1. Introduction

Given a sequence of $n \in \mathbb{N}$ independent realizations $\{X_j\}_{j=1}^n$ of the random variable X, having density f on $\mathbb{R}$, the Rosenblatt–Parzen kernel estimator (Rosenblatt, 1956; Parzen, 1962) of f is given by

$$f_n(x) = \frac{1}{n}\sum_{j=1}^{n}(S_{h_n}K)(x - X_j), \quad (1.1)$$

where S_{h_n} is an operator defined by

$$(S_{h_n}K)(x) = \frac{1}{h_n}K\left(\frac{x}{h_n}\right), \quad (1.2)$$

K is a kernel, i.e., a function on $\mathbb{R}$ such that $\int K(x)dx = 1$ and $h_n > 0$ is a non-stochastic bandwidth such that $h_n \to 0$ as $n \to \infty$.[1]

☆ We thank an anonymous referee and an Associate Editor for excellent comments that improved this note significantly.

* Corresponding author at: Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA.
E-mail addresses: kairat_mynbayev@yahoo.com (K. Mynbaev), carlos.martins@colorado.edu, c.martins-filho@cgiar.org (C. Martins-Filho).

1 Throughout this note, integrals are over $\mathbb{R}$, unless otherwise specified.

http://dx.doi.org/10.1016/j.spl.2016.11.019
0167-7152/

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019

One of the most natural and mathematically sound (Devroye and Györfi, 1985; Devroye, 1987) criteria to measure the performance of f_n as an estimator of f is the L_1 distance $\int |f_n - f|$. In particular, given that this distance is a random variable (measurable function of $\{X_j\}_{j=1}^n$) it is convenient to focus on $E\left(\int |f_n - f|\right)$, where E denotes the expectation taken using f. For this criterion, there is a simple bound (Devroye, 1987, p. 31)

$$E\left(\int |f_n - f|\right) \le \int |(f * S_{h_n}K) - f| + E\left(\int |f_n - f * S_{h_n}K|\right),$$

where for arbitrary $f, g \in L_1$, $(f*g)(x) = \int g(y)f(x-y)dy$ is the convolution of f and g. The term $\int |f * S_{h_n}K - f|$ is called bias over $\mathbb{R}$ and $E\left(\int |f_n - f * S_{h_n}K|\right)$ is called the variation over $\mathbb{R}$. There exists a large literature devoted to establishing conditions on f and K that assure suitable rates of convergence of the bias to zero as $n \to \infty$ (see, *inter alia*, Silverman, 1986; Devroye, 1987; Tsybakov, 2009). In particular, if K is of order s, i.e., $\alpha_j(K) = 0$ for $j = 1, \ldots, s-1$ and $\alpha_s(K) \neq 0$, where $\alpha_j(K) = \int t^j K(t)dt$ is the jth moment of K, and f has an integrable derivative $f^{(s)}$, then $\int |f * S_{h_n}K - f|$ is of order $O(h_n^s)$ and this order cannot be improved, see, e.g., Devroye (1987, Theorem 7.2). In this note, we show that if in (1.2) the kernel is allowed to depend on n, then the order $O(h_n^s)$ can be replaced by the order $o(h_n^s)$, without increasing the order of the kernel or the smoothness of the density. In addition, another result from Devroye (1987) states that if K is a kernel of order greater than s and the derivative $f^{(s)}$ is a-Lipschitz then the bias is of order $O(h_n^{s+a})$. We achieve the same rate of convergence with kernels of order s.

2. Main results

Let L_1 and C denote the spaces of integrable and (bounded) continuous functions on $\mathbb{R}$ with norms $\|f\|_1 = \int |f|$ and $\|f\|_C = \sup |f|$, and $\beta_s(K) = \int |t|^s \, |K(t)| \, dt$. Let $\{K_n\}$ be a sequence of kernels and define

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}(S_{h_n}K_n)(x - X_j).$$

In the following Theorem 1, the density f has the same degree of smoothness and the kernels K_n are of the same order as in Devroye (1987, Theorem 7.2), but the bias is of order $o(h_n^s)$ instead of $O(h_n^s)$. This results because the kernels depend on n and have "disappearing" moments of order s.

Theorem 1. *Let $\{K_n\}$ be a sequence of kernels of order s such that: 1. $\alpha_s(K_n) \to 0$; 2. $\{u^s K_n(u)\}$ is uniformly integrable. For all f with absolutely continuous $f^{(s-1)}$ and $f^{(s)} \in L_1$, we have $\|f * S_{h_n}K_n - f\|_1 = o(h_n^s)$.*

Proof. Note that since K_n is a kernel

$$f * S_{h_n}K_n(x) - f(x) = \int K_n(t)[f(x - h_n t) - f(x)]dt. \tag{2.1}$$

Since f is s-times differentiable, by Taylor's Theorem,

$$f(x - h_n t) - f(x) = \sum_{j=1}^{s-1}\frac{f^{(j)}(x)}{j!}(-h_n t)^j + \int_x^{x-h_n t}\frac{(x - h_n t - u)^{s-1}}{(s-1)!}f^{(s)}(u)du.$$

Furthermore, given that K_n is of order s,

$$f * S_{h_n}K_n(x) - f(x) = \frac{1}{(s-1)!}\iint_x^{x-h_n t}(x - h_n t - u)^{s-1}f^{(s)}(u)duK_n(t)dt. \tag{2.2}$$

Letting $\lambda = -\frac{u-x}{h_n t}$ we have

$$\int_x^{x-h_n t}(x - h_n t - u)^{s-1}f^{(s)}(u)du = (-h_n t)^s \int_0^1 f^{(s)}(x - h_n\lambda t)(1-\lambda)^{s-1}d\lambda. \tag{2.3}$$

Substituting (2.3) into (2.2) we obtain

$$f * S_{h_n}K_n(x) - f(x) = \frac{(-h_n)^s}{s!}\iint_0^1 f^{(s)}(x - h_n\lambda t)s(1-\lambda)^{s-1}d\lambda t^s K_n(t)dt. \tag{2.4}$$

Since $\int_0^1 (1-\lambda)^{s-1}d\lambda = \frac{1}{s}$, we have that

$$\frac{(-h_n)^s}{(s-1)!}\iint_0^1 f^{(s)}(x)(1-\lambda)^{s-1}d\lambda t^s K_n(t)dt = \frac{(-h_n)^s}{s!}f^{(s)}(x)\int t^s K_n(t)dt. \tag{2.5}$$

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019

Then, adding and subtracting (2.5) to the right-hand side of (2.4) gives

$$f * S_{h_n}K_n(x) - f(x) = \frac{(-h_n)^s}{s!}\left(f^{(s)}(x)\alpha_s(K_n) + \iint_0^1 [f^{(s)}(x - h_n\lambda t) - f^{(s)}(x)]s(1-\lambda)^{s-1}d\lambda t^s K_n(t)dt\right).$$

Since $f^{(s)} \in L_1$ we write its continuity modulus as $\omega(\delta) = \sup_{|t|\le\delta}\int \left|f^{(s)}(x-t) - f^{(s)}(x)\right| dx$. It is well-known (see properties M.2, M.6 and M.7 in Zhuk and Natanson (2003)) that

$$\omega(\delta) \le 2\left\|f^{(s)}\right\|_1, \quad \omega \text{ is nondecreasing and } \lim_{\delta\to 0}\omega(\delta) = 0. \tag{2.6}$$

Then,

$$\begin{aligned}
\|f * S_{h_n}K_n - f\|_1 &\le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + \iint_0^1 \int \left|f^{(s)}(x - h_n\lambda t) - f^{(s)}(x)\right| dx\, s(1-\lambda)^{s-1}d\lambda |t^sK_n(t)|dt\right] \\
&\le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + \int_0^1\int \omega(\lambda h_n|t|)\left|t^sK_n(t)\right| dt\, s(1-\lambda)^{s-1}d\lambda\right] \\
&= \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + \int_0^1\left(\int_{|t|\le\frac{1}{\sqrt{h_n}}} + \int_{|t|>\frac{1}{\sqrt{h_n}}}\right)\omega(\lambda h_n|t|)|t^sK_n(t)|dt\, s(1-\lambda)^{s-1}d\lambda\right] \quad (2.7) \\
&\le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + \omega\left(\sqrt{h_n}\right)\beta_s(K_n) + 2\left\|f^{(s)}\right\|_1\int_{|t|>\frac{1}{\sqrt{h_n}}}\left|t^sK_n(t)\right| dt\right]. \quad (2.8)
\end{aligned}$$

Given that $\alpha_s(K_n) \to 0$ as $n \to \infty$, $\{t^sK_n(t)\}$ is uniformly integrable, which implies $\sup_n \beta_s(K_n) < \infty$, and using (2.6) and (2.8) we have

$$\|f * S_{h_n}K_n - f\|_1 = o(h_n^s). \quad \square \tag{2.9}$$

Remark 1. Kernel sequences $\{K_n\}$ that satisfy the restrictions imposed by Theorem 1 can be easily constructed. To this end, denote by $\mathcal{B}_s$ the space of functions with bounded norm $\|K\|_{\mathcal{B}_s} = \beta_0(K) + \beta_s(K)$. Take functions $K_{(0)}, K_{(s)} \in \mathcal{B}_s$ such that

$$\alpha_0(K_{(0)}) = 1, \qquad \alpha_j(K_{(0)}) = 0 \quad \text{for } j = 1, \ldots, s; \qquad \alpha_j(K_{(s)}) = 0 \quad \text{for } j = 0, 1, \ldots, s-1, \qquad \alpha_s(K_{(s)}) = 1. \tag{2.10}$$

We define the n-dependent kernel $K_n = K_{(0)} + h_nK_{(s)}$ with $0 < h_n \le 1$. Note that K_n is a kernel of order s with $\alpha_s(K_n) = h_n$ which tends to zero as $n \to \infty$. It is clear that any kernel K of order s can be written as $K = K_{(0)} + \alpha_s(K)K_{(s)}$, so that the conventional s-order kernels obtain from ours with $\alpha_s(K) = h_n$. Furthermore, it follows from (2.10) that $\{t^s(K_{(0)}(t) + h_nK_{(s)}(t))\}$ is uniformly integrable.

Now, to obtain $K_{(0)}$ and $K_{(s)}$, assume that for a nonnegative kernel K we have $\beta_{2s}(K) < \infty$. Then, we can associate with K a symmetric matrix

$$A_s = \begin{pmatrix} \alpha_0(K) & \alpha_1(K) & \cdots & \alpha_s(K) \\ \alpha_1(K) & \alpha_2(K) & \cdots & \alpha_{s+1}(K) \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_s(K) & \alpha_{s+1}(K) & \cdots & \alpha_{2s}(K) \end{pmatrix},$$

such that $\det A_s \ne 0$ (see Mynbaev et al., 2014). For an arbitrary vector $b \in \mathbb{R}^{s+1}$ let $a = A_s^{-1}b$ and define a polynomial transformation of K by $(T_aK)(t) = \left(\sum_{i=0}^s a_it^i\right)K(t)$. Then, we put $K_{(0)} = T_aK$ with $b = (1, 0, \ldots, 0)'$ and $K_{(s)} = T_aK$ with $b = (0, \ldots, 0, 1)'$, which satisfy (2.10). Thus, we have the following corollary to Theorem 1.

Corollary 1. *Let $K_n = K_{(0)} + h_nK_{(s)}$ where $K_{(0)}$ and $K_{(s)}$ are as defined in Remark 1. Then, for all f with absolutely continuous $f^{(s-1)}$ and $f^{(s)} \in L_1$, we have $\|f * S_{h_n}K_n - f\|_1 = o(h_n^s)$.*

Remark 2. If K_n are supported on $[-M, M]$ for some $M > 0$ and for all n, then in (2.7), instead of splitting $\mathbb{R} = \left\{|t| \le 1/\sqrt{h_n}\right\} \cup \left\{|t| > 1/\sqrt{h_n}\right\}$ we can use $\mathbb{R} = \{|t| \le M\} \cup \{|t| > M\}$ and then instead of (2.8) we get

$$\left\|f * S_{h_n}K_n - f\right\|_1 \le \frac{h_n^s}{s!}\left[\left\|f^{(m)}\right\|_1|\alpha_s(K_n)| + \omega(h_nM)\beta_s(K_n)\right].$$

Hence, selecting $\{K_n\}$ in such a way that $\alpha_s(K_n) = O(\omega(h_n))$, $\sup_n \beta_s(K_n) < \infty$ and using the fact that $\omega(h_nM) \le (M+1)\,\omega(h_n)$[2] we get a result that is more precise than (2.9), i.e.,

$$\left\|f * S_{h_n}K_n - f\right\|_1 = O(h_n^s\omega(h_n)).$$

[2] See property M.5 in Zhuk and Natanson (2003).

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019

Remark 3. By Young's inequality, the variation of $\hat{f}_n$ using $K_n = K_{(0)} + h_n K_{(s)}$ is such that

$$E\int|\hat{f}_n - f * S_{h_n}K_n| \le E\int|\hat{f}_n - f * S_{h_n}K_{(0)}| + h_n\int|f * S_{h_n}K_{(s)}|$$
$$\le E\int|\hat{f}_n - f * S_{h_n}K_{(0)}| + h_n\int|f|\int|K_{(s)}|.$$

Letting $f_n^{(0)}$ be the estimator in (1.1) with $K = K_{(0)}$, we have $E\int|\hat{f}_n - f * S_{h_n}K_{(0)}| \le E\int|f_n^{(0)} - f * S_{h_n}K_{(0)}| + h_n\int|f|\int|K_{(s)}|$. Hence,

$$E\int|\hat{f}_n - f * S_{h_n}K_n| \le E\int|f_n^{(0)} - f * S_{h_n}K_{(0)}| + 2\,h_n\int|f|\int|K_{(s)}|.$$

Since, $h_n \to 0$ as $n \to \infty$, the variation of $\hat{f}_n$ is asymptotically bounded by the variation of the conventional estimator f_n using $K_{(0)}$, i.e., $E\int|f_n^{(0)} - f * S_{h_n}K_{(0)}|$.

Under the assumptions that f has a variance, $\int(1+t^2)(K_{(0)}(t))^2dt < \infty$, Devroye (1987, Theorem 7.4) showed that $E\int|f_n^{(0)} - f * S_{h_n}K_{(0)}| = O((nh_n)^{-1/2})$. Thus,

$$\sqrt{nh_n}E\int|\hat{f}_n - f * S_{h_n}K_n| = (1 + (nh_n^3)^{1/2})O(1) = O(1),$$

where the last equality follows if $nh_n^3 \le c < \infty$.

We now provide an analog for Theorem 7.1 in Devroye (1987). There, the bias order $O(h^{s+a})$ is achieved for kernels with orders greater than s, while in the following theorems we obtain the same order of bias for kernels of order s.

Theorem 2. *Let $\{K_n\}$ be a sequence of kernels of order s such that: 1. $\alpha_s(K_n) = O(h_n^a)$; 2. $\sup_n \beta_{s+a}(K_n) < \infty$, for some $a \in (0, 1]$. For f with absolutely continuous $f^{(s-1)}$ and $f^{(s)} \in L_1$ assume that for some $0 < c < \infty$*

$$\omega(\delta) \le c\,|\delta|^a. \quad (2.11)$$

Then, $\|f * S_{h_n}K_n - f\|_1 = O(h_n^{s+a})$.

Proof. As in the proof of Theorem 1, we have

$$\|f * S_{h_n}K_n - f\|_1 \le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + \int_0^1\int\omega(\lambda h_n|t|)s(1-\lambda)^{s-1}\left|t^sK_n(t)\right|dtd\lambda\right]. \quad (2.12)$$

By (2.11)

$$\|f * S_{h_n}K_n - f\|_1 \le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1|\alpha_s(K_n)| + ch_n^a\int\int_0^1 s(1-\lambda)^{s-1}d\lambda\left|t^{s+a}K_n(t)\right|dt\right].$$

Hence, under conditions 1 and 2 in the statement of the theorem we have $\|f * S_{h_n}K_n - f\|_1 = O(h_n^{s+a})$. □

Remark 4. Practitioners may find condition (2.11) too general, preferring more primitive conditions on $f^{(s)}$. To this end, we say that a function g defined on $\mathbb{R}$ satisfies a *global Lipschitz condition* of order $a \in (0, 1]$ if there exist positive functions $l(x), r(x)$ such that

$$|g(x-h) - g(x)| \le l(x)|h|^a \quad \text{for } |h| \le r(x),\ x \in \mathbb{R}. \quad (2.13)$$

The function l is called a Lipschitz constant and the function r is called a Lipschitz radius. The class $Lip(a, \delta)$, for $\delta > 1$, is defined as the set of functions g which satisfy (2.13) with l and r such that

$$\int(l(x) + r(x)^{-\delta})dx < \infty. \quad (2.14)$$

In the next lemma we give two sufficient sets of conditions for $g \in Lip(a, \delta)$. In the first case g is compactly supported, and in the second it is not.

Lemma 1. (a) *Suppose g has compact support,* $\operatorname{supp} g \subseteq [-N, N]$ *for some $N > 0$, and g satisfies the usual Lipschitz condition $|g(x-h) - g(x)| \le c|h|^a$ for any x, h and some $a \in (0, 1]$. Set $l(x) = c,\ r(x) = 1$ for $|x| < N$ and $l(x) = 0,\ r(x) = |x| - N$ for $|x| \ge N$. Then, $g \in Lip(a, \delta)$ with any $\delta > 1$.*

(b) *Suppose that $|g^{(1)}(t)| \le ce^{-|t|},\ t \in \mathbb{R}$. Let $l(x) = c\exp(-|x|/2 + 1/2),\ r(x) = (1 + |x|)/2,\ x \in \mathbb{R}$. Then, $g \in Lip(1, \delta)$ with any $\delta > 1$.*

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019

Proof. (a) If $|x| < N$, then $|g(x-h)-g(x)| \le |h|^a l(x)$ for all h (and not only for $|h| \le r(x)$). If $|x| \ge N$, then $|h| \le r(x) = |x|-N$ implies $|x-h| \ge |x|-|h| \ge N$, so that $|g(x-h)-g(x)| = 0 = |h|^a l(x)$ for $|h| \le r(x)$.

(b) Let $|x| \ge 1$. We have

$$|g(x-h)-g(x)| \le |h| \sup_{|t-x|\le |h|} |g^{(1)}(t)| \le |h| c \sup_{|t-x|\le|h|} e^{-|t|}. \tag{2.15}$$

$|t-x| \le |h| \le r(x) = (1+|x|)/2$ implies $|t| = |x+t-x| \ge |x|-|t-x| \ge |x|/2-1/2 \ge 0$ and (2.15) gives $|g(x-h)-g(x)| \le |h| c \exp(1/2-|x|/2) = |h| l(x)$ for $|h| \le r(x)$. Now let $|x| < 1$. Then, $e^{-|x|/2} \ge e^{-1/2}$ so that by (2.15), $|g(x-h)-g(x)| \le |h|c \le |h| c \exp(1/2-|x|/2) = |h| l(x)$ for all h and not only for $|h| \le r(x)$. Condition (2.14) is obviously satisfied in both cases. □

By part (a) of Lemma 1, compactly supported densities with derivative $f^{(s)}$ that satisfies the usual a-Lipschitz condition are such that $f^{(s)} \in Lip(a,\delta)$ for any $\delta > 1$. This corresponds to the case treated in Theorem 7.1 of Devroye (1987). Part (b) shows that for densities with unbounded domains, not covered by Theorem 7.1, if $f^{(s)}(x)$ has derivative that decays exponentially as $|x| \to \infty$, then $f^{(s)} \in Lip(1,\delta)$ for any $\delta > 1$. Next we provide a version of Theorem 2 for densities with derivative $f^{(s)} \in Lip(a,\delta)$.

Theorem 3. *Suppose that the density f is such that its derivative $f^{(s)}$ belongs to L_1, C (the respective norms are finite) and $Lip(a,\delta)$. Let $\{K_n\}$ be a sequence of kernels of order s such that: 1. $\alpha_s(K_n) = O(h_n^a)$; 2. $\sup_n \max\{\beta_{s+a}(K_n), \beta_{s+\delta}(K_n)\} < \infty$. Then, $\|f * S_{h_n}K_n - f\|_1 = O(h_n^{s+a})$.*

Proof. As in the proof of Theorem 1, we have

$$\|f * S_{h_n}K_n - f\|_1 \le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1 |\alpha_s(K_n)| + \int_0^1 \iint \left|f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)\right| |t^s K_n(t)| dt\, dx\, s(1-\lambda)^{s-1} d\lambda\right].$$

Let $\mathcal{I}(x) = \int \left|f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)\right| |t^s K_n(t)| dt$ and note that since $f^{(s)} \in Lip(a,\delta)$

$$\begin{aligned}\mathcal{I}(x) &= \left(\int_{\lambda h_n|t|\le r(x)} + \int_{\lambda h_n|t|>r(x)}\right) \left|f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)\right| |t^s K_n(t)| dt \\ &\le \int_{\lambda h_n|t|\le r(x)} l(x)\lambda^a h_n^a |t|^{a+s} |K_n(t)| dt + \int_{\lambda h_n|t|>r(x)} |f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)||t^s K_n(t)| dt \\ &\le h_n^a \beta_{s+a}(K_n) l(x) + \int_{\lambda h_n|t|>r(x)} |f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)||t^s K_n(t)| dt.\end{aligned}$$

Letting $\mathcal{I}_1(x) = \int_{\lambda h_n|t|>r(x)} |f^{(s)}(x-h_n\lambda t) - f^{(s)}(x)||t^s K_n(t)| dt$ we have

$$\mathcal{I}_1(x) \le \int_{\lambda h_n|t|>r(x)} \frac{1}{|t|^\delta}\left(|f^{(s)}(x-h_n\lambda t)| + |f^{(s)}(x)|\right) |t|^{s+\delta} |K_n(t)| dt.$$

Noting that $|t|^{-\delta} < \lambda^\delta h_n^\delta r(x)^{-\delta}$ and given that $\|f^{(s)}\|_C < \infty$, we obtain $\mathcal{I}_1(x) \le 2\|f^{(s)}\|_C \frac{\lambda^\delta h_n^\delta}{r(x)^\delta} \beta_{s+\delta}(K_n)$. Consequently,

$$\mathcal{I}(x) \le h_n^a \max\{\beta_{s+a}(K_n), \beta_{s+\delta}(K_n)\}\left(l(x) + 2h_n^{\delta-a}\|f^{(s)}\|_C \frac{1}{r(x)^\delta}\right). \tag{2.16}$$

Since $\int_0^1 s(1-\lambda)^{s-1} ds = \frac{1}{s}$ and given (2.14)

$$\|f * S_{h_n}K_n - f\|_1 \le \frac{h_n^s}{s!}\left[\left\|f^{(s)}\right\|_1 |\alpha_s(K_n)| + \frac{1}{s} h_n^a \max\{\beta_{s+a}(K_n), \beta_{s+\delta}(K_n)\} \int \left(l(x) + 2\|f^{(s)}\|_C \frac{1}{r(x)^\delta}\right) dx\right]. \tag{2.17}$$

Thus, using conditions 1. and 2. in the statement of the theorem, we have $\|f * S_{h_n}K_n - f\|_1 = O(h_n^{s+a})$. □

Remark 5. As in the case of Theorems 1–3 do not address the construction of the kernel sequence $\{K_n\}$. The following corollary to Theorem 3 shows that $K_n = K_{(0)} + h_n^a K_{(s)}$ is a suitable kernel sequence, where $K_{(0)}$ and $K_{(s)}$ are as defined above.

Corollary 2. *Suppose the density f is such that its derivative $f^{(s)}$ belongs to L_1, C and to $Lip(a,\delta)$, where $a \in (0,1]$, $\delta > 1$. Let $K_{(0)}, K_{(s)}$ satisfy (2.10) and belong to the intersection $\mathcal{B}_{s+a} \cap \mathcal{B}_{s+\delta}$. Put $K_n = K_{(0)} + h_n^a K_{(s)}$, $0 \le h_n \le 1$. Then K_n is a kernel of order s for $h_n > 0$ and $\|f * S_{h_n}K_n - f\|_1 = O(h_n^{s+a})$.*

The condition $K_{(0)} \in \mathcal{B}_{s+a}$ and the definition $K_n = K_{(0)} + h_n^a K_{(s)}$ can be replaced by $K_{(0)} \in \mathcal{B}_{s+1}$ and $K_n = K_{(0)} + h_n K_{(s)}$, respectively, without affecting the conclusion.

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019

References

Devroye, L., 1987. A Course in Density Estimation. Birkhäuser, Boston, MA.
Devroye, L., Györfi, L., 1985. Nonparametric Density Estimation: The L_1 View. John Wiley and Sons, New York, NY.
Mynbaev, K., Nadarajah, S., Withers, C., Aipenova, A., 2014. Improving bias in kernel density estimation. Statist. Probab. Lett. 94, 106–112.
Parzen, E., 1962. On estimation of a probability density and mode. Ann. Math. Stat. 33, 1065–1076.
Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. 27, 832–837.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.
Tsybakov, A.B., 2009. Introduction to Nonparametric Estimation. Springer-Verlag, New York, NY.
Zhuk, V.V., Natanson, G.I., 2003. Seminorms and continuity modules of functions defined on a segment. J. Math. Sci. 118, 4822–4851.

Please cite this article in press as: Mynbaev, K., Martins-Filho, C., Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view. Statistics and Probability Letters (2016), http://dx.doi.org/10.1016/j.spl.2016.11.019