УДК 519.221

Aipenova A.S.
Al-Farabi Kazakh National University, Almaty, Kazakhstan
E-mail: a.aipenova@mail.ru

Some asymptotic results for kernel density estimation with Lipschitz smoothness

Estimates of density derivatives can be used to evaluate modes and inflection points of f and can be applied, for example, to the bandwidth choice for the estimation of densities themselves. In this paper we generalized the nonparametric kernel density estimator suggested by [6] to the estimation of 1^{st} order density derivatives. Some results for this estimator are obtained. Our results are based on imposing global Lipschitz conditions on f and applying the kernel suggested by [6]. An integral representation for the bias and the exact orders of the bias and variance of the estimator are obtained. The conditions of consistency and uniform consistency of this estimator are studied. A criterion for the optimal bandwidth that minimizes asymptotic integrated mean squared error is provided. The general case will be considered elsewhere.
Key words: kernel density estimation, Lipschitz smoothness, density derivative estimator, uniform consistency, bandwidth choice.

А.С. Аипенова

Некоторые асимптотические результаты для оценки плотности ядра с Липшицевой гладкости

Оценки производных плотности могут быть использованы для оценок мод, точек перегиба f, а также, к примеру, для выбора полосы пропускания оценки самих плотностей. В данной работе мы обобщаем непараметрические ядерные оценки плотности, предложенные в [6], на случай оценки производной плотности первого порядка. Для этой оценки получены некоторые результаты. Наши результаты основаны на введении глобальных условий Липшица на f и применении ядра предложенного в [6]. Получено интегральное представление для смещения, получены точные по порядку оценки для смещения и вариации оценки. Условия согласованности и равномерной согласованности этой оценки изучены. Найден критерий для оптимальной полосы пропускания, который минимизирует асимптотическую интегрированную среднеквадратичную ошибку. Общий случай будет рассмотрен в другой статье.
Ключевые слова: оценка плотности ядра, Липшицевая гладкость, оценка производных плотности, равномерная согласованность, выбор полосы пропускания.

А.С. Аипенова

Липшицтік тегістік әдісімен тығыздық ядросын бағалау үшін кейбір асимптотикалық нәтижелер

Модтарды бағалауға, f-тің иілу нүктелерін табуға және де тығыздықтардың өздерін бағалау үшін қадамды таңдауға тығыздық туындыларының бағалаулары қолданылады. Бұл жұмыста [6]-да келтірілген тығыздықтың параметрлік емес ядролық бағалауларын тығыздықтың бірінші ретті туындысының бағалау жағдайына жалпыладық. Осы бағалауға кейбір нәтижелер алынды. Біздің нәтижелеріміз глобальды Липшиц шарттарын f-ке енгізуге және [6]-да келтірілген ядроны қолдануға негізделген. Осы бағалаудың ығысуына интегральдық өрнек және ығысуы мен вариациясының нақты реті алынды. Бұл бағалаудың сәйкестілік шарты және бірқалыпты сәйкестілігі зерттелді. Тиімді қадам критерийі асимптотикалық интегралданған орташа квадраттық қателікті азайту арқылы табылды. Жалпы жағдай басқа мақалада қарастырылады.
Түйін сөздер: тығыздық ядросың бағалауы, Липшицтік тегістік, тығыздық туындыларының бағалауы, бірқалыпты сәйкестілік, қадамды таңдау.

Introduction

Let $X_1, X_2, ..., X_n$ be independent identically distributed random variables with common probability density function f. Let f' denote the 1^{st} order derivative of f. Kernel estimator is a nonparametric way to estimate the probability density function of a random variable.

The Rosenblatt-Parzen estimator for the density f evaluated at $x \in \Re$ is defined by

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

where h_n is a sequence of positive numbers converging to 0 and K is a kernel on $\Re$ satisfying

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

A symmetric kernel function satisfies $K(t) = K(-t)$ for all t.

Various properties of these estimates including uniform consistency are well known, see for example [1] and [2]. Estimation of a probability density function and its derivatives have been considered by several authors, including [3], [4] and [5]. [5] studied asymptotic properties of density estimates and its derivatives using the kernel method. [4] showed that the uniform continuity of f was necessary for uniform consistency, under the condition $\sum \exp(-cnh^2) < \infty$. This condition is substantially weakened in [5]. [7] estimated the first derivative when the density is a mixture of univariate exponential densities with respect to Lebesgue measure. They also investigated the consistency and the mean squared error convergence properties of these estimates.

In this paper we generalize the nonparametric kernel density estimator suggested by [6] to the estimation of 1^{st} order density derivatives. We provide asymptotic characterization of the proposed estimator, including uniform consistency. In addition, we discuss optimal bandwidth selection based on the minimization of an asymptotic approximation for the integrated mean squared error. The material of section 2 repeats the necessary definitions of

[6, p. 220-222]. Section 3 provides some asymptotic results of new kernel density derivative estimation and derives the expression for bias. Section 4 provides asymptotic property of the proposed estimator and discusses orders of the bias and variance. Section 5 discusses a global criterion of the choice of bandwidth.

Lipschitz conditions and kernels

The properties of nonparametric density estimators are traditionally obtained by assumptions on the smoothness of the underlying density. 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 $h \in \Re$. Here, we focus on centered even-order differences because the resulting kernels are symmetric. Let $C_{2k}^{l} = \frac{(2k)!}{(2k-l)!l!}, l = 0, ..., 2k, k \in \aleph$ be the binomial coefficients, $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$, $s = -k, ..., k$ and

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

They say that a function $f : \Re \to \Re$ satisfies the Lipschitz condition of order $2k$ if for any $x \in \Re$ there exist $H(x) > 0$ and $\varepsilon(x) > 0$ such that $\left|\Delta_h^{2k} f(x)\right| \leq H(x) h^{2k}$ for all h such that $|h| \leq \varepsilon(x)$. They call $H(x)$ a Lipschitz constant and $\varepsilon(x)$ a Lipschitz radius. For a kernel K [6] define a new set of kernels $\{M_k(x)\}_{k=1,2,3,\cdots}$ where

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

In their context [6] K is a seed kernel for M_k. The main impetus for the definition of $M_k(x)$ is that it allows them to express the bias of their proposed estimator in terms of higher order finite differences of the density derivative $f(x)$.

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

Let $\lambda_{k,s} = \frac{(-1)^{s+1}(k!)^2}{(k+s)!(k-s)!}, s = 1, ..., k$ and since $-\frac{c_{k,s}}{c_{k,0}} = -\frac{c_{k,-s}}{c_{k,0}} = \lambda_{k,s}, s = 1, ..., k$, (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 $M_k(x) = M_k(-x), x \in \Re$.

Since the coefficients $c_{k,s}$ satisfy $\sum_{|s|=0}^{k} c_{k,s} = (1-1)^{2k} = 0$, we have

$$-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} = 1 \text{ or } \sum_{s=1}^{k} \lambda_{k,s} = \frac{1}{2}.$$

Consequently, (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 $\{M_k(x)\}_{k=1,2,\cdots}$ is a class of kernels. [8] 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.

New nonparametric kernel density estimator

In this section, we take the 1^{st} order derivative of the [6]'s estimator to define the derivative estimator.

$$\hat{f}'_k(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^2}M'_k\left(\frac{x-X_j}{h_n}\right) = \frac{1}{n}\sum_{j=1}^{n}u_j \tag{4}$$

where $u_j = \frac{1}{h_n^2}M'_k\left(\frac{x-X_j}{h_n}\right)$ and

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

Given the independent and identically distributed (IID) assumption (maintained everywhere), we have

$$E\hat{f}'_k(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n^2}E\left(M'_k\left(\frac{x-X_j}{h_n}\right)\right) = \frac{1}{n}\sum_{j=1}^{n}E(u_j) = Eu_1 \tag{6}$$

and

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

At first, we need to impose restrictions on K and f needed to obtain a suitable representation for the bias and variance of the density derivative estimators. Hence, we assume that

Assumption 1.

a) $|K(s)| = o\left(\frac{1}{|s|}\right), |s| \to \infty$

b) $|f(s)| = O(s), |s| \to \infty$.

Assumption 1 is used to obtain an integral representation for the bias $B(\hat{f}'_k(x)) = E(\hat{f}'_k(x)) - f'(x)$ of $\hat{f}'_k(x)$ in terms of centered even order differences of $f'(x)$.

Theorem 1 *Under Assumption 1, for any $h_n > 0$, $B(\hat{f}'_k(x)) = -\frac{1}{c_{k,0}}\int\limits_{-\infty}^{+\infty}K(t)\Delta_{h_nt}^{2k}f'(x)dt$.*

Proof. Under Assumption 1, we have that $\left|K(\frac{l}{s})f(x-h_nl)\right| = |\frac{l}{s}=t| = |K(t)f(x-sh_nt)| = o\left(\frac{1}{|t|}\right)O(|x-sh_nt|) = o\left(\left|\frac{x-sh_nt}{t}\right|\right) = o(1)$, as $|t|\to\infty$, $h_n>0$.

Therefore by (6), apply the change of variables, by (5) and we can integrate by parts, we get

$$E(\hat{f}'_k(x)) = E(u_1) = \frac{1}{h_n^2}\int\limits_{-\infty}^{+\infty} M'_k\left(\frac{x-t}{h_n}\right) f(t)dt =$$

$$= \frac{1}{h_n}\int\limits_{-\infty}^{+\infty} M'_k(l) f(x-h_n l)dl = -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|sh_n}\int\limits_{-\infty}^{+\infty} K'\left(\frac{l}{s}\right) f(x-h_n l)dl =$$

$$= -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|s}\left[\frac{1}{h_n}K\left(\frac{l}{s}\right) f(x-h_n l)|_{-\infty}^{+\infty} + s\int\limits_{-\infty}^{+\infty} K\left(\frac{l}{s}\right) f'(x-h_n l)dl\right] =$$

$$= -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\int\limits_{-\infty}^{+\infty} K\left(\frac{l}{s}\right) f'(x-h_n l)dl = \tag{8}$$

$$= -\frac{1}{c_{k,0}}\left[\sum_{s=-k}^{-1}\frac{c_{k,s}}{-s}(-s)\int\limits_{-\infty}^{+\infty} K(-t)f'(x+sh_n t)dt + \right.$$

$$\left. + \sum_{s=1}^{k}\frac{c_{k,s}}{s}(-s)\int\limits_{+\infty}^{-\infty} K(-t)f'(x+sh_n t)dt\right] =$$

$$= -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k} c_{k,s}\int\limits_{\Re} K(-t)f'(x+sh_n t)dt$$

Hence, from (8), (2) and (1) we obtain

$$B(\hat{f}'(x)) = -\frac{1}{c_{k,0}}\left[\sum_{s=-k}^{-1} c_{k,s}\int\limits_{-\infty}^{+\infty} K(-t)f'(x+sh_n t)dt + \right.$$

$$\left. + \sum_{s=1}^{k} c_{k,s}\int\limits_{-\infty}^{+\infty} K(-t)f'(x+sh_n t)dt\right] - \frac{c_{k,0}}{c_{k,0}}f'(x) =$$

$$= -\frac{1}{c_{k,0}}\left[\sum_{|s|=1}^{k} c_{k,s}\int\limits_{-\infty}^{+\infty} K(-t)f'(x+sh_n t)dt + c_{k,0}\int\limits_{-\infty}^{+\infty} K(-t)f'(x+0h_n t)dt\right] = \tag{9}$$

$$= -\frac{1}{c_{k,0}}\int\limits_{-\infty}^{+\infty} K(-t)\sum_{|s|=0}^{k} c_{k,s}f'(x+sh_n t)dt = -\frac{1}{c_{k,0}}\int\limits_{-\infty}^{+\infty} K(-t)\Delta_{h_n t}^{2k} f'(x)dt.$$

The theorem has been proved.

Asymptotic property of $\hat{f}'_k(x)$ and orders of the bias and variance

In this section we give an asymptotic characterization of the estimator. Consistency of $\hat{f}'_k(x)$ is provided by the following theorem.

Theorem 2 *Assume that: a) The characteristic function ϕ_K of K satisfies $\int\limits_{\Re} |s\phi_K(s)|ds < \infty$; b) $f_k'(x)$ is bounded and uniformly continuous in $\Re$; c) $nh_n^4 \to \infty$ as $n \to \infty$. Then $\hat{f}_k'(x)$ is uniformly consistent, that is,*

$$plim_{n\to\infty} \sup_{x\in\Re} \left(|\hat{f}_k'(x) - f'(x)| \right) = 0.$$

Proof. To establish the uniform consistency $\hat{f}_k'(x)$, we denote $\psi_j = \frac{x-X_j}{h}$, then (4) yields $\hat{f}_k'(x) = \frac{1}{n}\sum\limits_{j=1}^{n} \frac{1}{h^2} M_k'(\psi_j)$ and using (5) we get

$$M_k'(\psi_j) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s} K'\left(\frac{\psi_j}{s}\right) \tag{10}$$

By a) the inversion theorem for Fourier transforms means that

$$K'\left(\frac{\psi_j}{s}\right) = \frac{(-i)}{2\pi} \int\limits_{\Re} \exp\left\{\frac{-it\psi_j}{s}\right\} t\phi_K(t)dt. \tag{11}$$

Using (4), (5), (11) and by changing variables in integration we have

$$\begin{aligned}
\hat{f}_k'(x) &= \frac{1}{nh_n^2}\sum_{j=1}^{n} M_k'\left(\frac{x-X_j}{h_n}\right) = \frac{1}{nh_n^2}\sum_{j=1}^{n} M_k'(\psi_j) = \\
&= \frac{i}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{nh_n^2} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s} \int\limits_{-\infty}^{+\infty} \exp\left\{\frac{-it\psi_j}{s}\right\} t\phi_K(t)dt = \\
&= \frac{i}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{nh_n^2} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|s} \int\limits_{-\infty}^{+\infty} \exp\left\{-it\left(\frac{x-X_j}{sh_n}\right)\right\} t\phi_K(t)dt = \\
&= \frac{i}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{n}\Bigg[\sum_{s=-k}^{-1} \frac{c_{k,s}}{-s} s \int\limits_{+\infty}^{-\infty} \exp\{-i\tau x\}\exp\{i\tau X_j\}\tau\phi_K(sh_n\tau)d\tau + \\
&+ \sum_{s=1}^{k} \frac{c_{k,s}}{s} s \int\limits_{-\infty}^{+\infty} \exp\{-i\tau x\}\exp\{i\tau X_j\}\tau\phi_K(sh_n\tau)d\tau \Bigg] = \\
&= \frac{i}{2\pi c_{k,0}} \sum_{j=1}^{n} \frac{1}{n} \sum_{|s|=1}^{k} c_{k,s} \int\limits_{-\infty}^{+\infty} \exp\{-i\tau x\}\exp\{i\tau X_j\}\tau\phi_K(sh_n\tau)d\tau = \\
&= \frac{i}{2\pi c_{k,0}} \int\limits_{-\infty}^{+\infty} \exp\{-i\tau x\} \sum_{j=1}^{n} \frac{1}{n}\exp\{i\tau X_j\} \sum_{|s|=1}^{k} c_{k,s}\tau\phi_K(sh_n\tau)d\tau = \\
&= \frac{i}{2\pi c_{k,0}} \int\limits_{-\infty}^{+\infty} \exp\{-i\tau x\}\hat{\phi}_n(\tau)\Delta(\tau)d\tau
\end{aligned} \tag{12}$$

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

$$E(\hat{f}'_k(x)) = \frac{i}{2\pi c_{k,0}} \int\limits_{\Re} \exp\{-i\tau x\} E\hat{\phi}_n(\tau)\Delta(\tau)d\tau = \frac{i}{2\pi c_{k,0}} \int\limits_{\Re} \exp\{-i\tau x\}\phi_f(\tau)\Delta(\tau)d\tau \quad (13)$$

Hence, by (12) and (13)

$$|\hat{f}'_k(x) - E(\hat{f}'_k(x))| \leq c \int\limits_{\Re} |\hat{\phi}_n(\tau) - \phi_f(\tau)||\exp\{-i\tau x\}||\Delta(\tau)|d\tau.$$

Since $|\exp\{-i\tau x\}| = 1$,

$$\sup_{x\in\Re} |\hat{f}'_k(x) - E(\hat{f}'_k(x))| \leq c \int\limits_{\Re} |\hat{\phi}_n(\tau) - \phi_f(\tau)||\Delta(\tau)|d\tau.$$

with no sup on the right-hand side because it does not depend on x. It follows from Lemma 2.1 of [7] that $\sup|\cdot|$ is measurable, its expectation is well defined and

$$E\left(\sup_{x\in\Re} |\hat{f}'_k(x) - E\hat{f}'_k(x)|\right) \leq c \int\limits_{\Re} E\left|\hat{\phi}_n(\tau) - \phi_f(\tau)\right| |\Delta(\tau)|d\tau.$$

Now,

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

$$= E(|Y_1 + iY_2|) = E|Y_1| + E|Y_2| \leq (EY_1^2)^{\frac{1}{2}} + (EY_2^2)^{\frac{1}{2}}$$

where

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

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

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

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

and

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

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

Consequently,

$$(V(\cos(\tau X_1)))^{\frac{1}{2}} \le \left[E\cos^2(\tau X_1) + (E\cos(\tau X_1))^2\right]^{\frac{1}{2}} \le$$

$$\le \left[E\cos^2(\tau X_1) + E\cos^2(\tau X_1)\right]^{\frac{1}{2}} \le \sqrt{2}$$

and

$$(V(\sin(\tau X_1)))^{\frac{1}{2}} \le \left[E\sin^2(\tau X_1) + (E\sin(\tau X_1))^2\right]^{\frac{1}{2}} \le$$

$$\le \left[E\sin^2(\tau X_1) + E\sin^2(\tau X_1)\right]^{\frac{1}{2}} \le \sqrt{2}$$

Hence, $(EY_1^2)^{\frac{1}{2}} + (EY_2^2)^{\frac{1}{2}} \le \frac{2\sqrt{2}}{\sqrt{n}}$. Then, $E\left(\left|\hat{\phi}_n(\tau) - \phi_f(\tau)\right|\right) \le \frac{2\sqrt{2}}{\sqrt{n}}$ and

$$\int\limits_{\Re} |\Delta(\tau)|d\tau \le \sum_{|s|=1}^{k} |c_{k,s}| \int\limits_{\Re} |\tau||\phi_K(sh_n\tau)|d\tau \le$$

$$\le \frac{1}{h_n^2}\sum_{|s|=1}^{k}\frac{|c_{k,s}|}{s^2}\int\limits_{\Re}|t\phi_K(t)|dt = \frac{c}{h_n^2}\int\limits_{\Re}|t\phi_K(t)|dt.$$

Finally,

$$E\left(\sup_{x\in\Re}|\hat{f}_k'(x) - E\hat{f}_k'(x)|\right) \le \frac{c}{h_n^2\sqrt{n}}\int\limits_{\Re}|t\phi_K(t)|dt$$

by condition c) tends to zero as $n \to \infty$. Furthermore, using Markov's inequality, we get

$$P(\sup_{x\in\Re}|\hat{f}_k'(x) - E\hat{f}_k'(x)| > \varepsilon) \to 0 \tag{14}$$

as $n \to \infty$, for all $\varepsilon > 0$, implying that $\sup_{x\in\Re}|\hat{f}_k'(x) - E\hat{f}_k'(x)| \xrightarrow{p} 0$. Finally,

$$\sup_{x\in\Re}|\hat{f}_k'(x) - f'(x)| \le \sup_{x\in\Re}|\hat{f}_k'(x) - E\hat{f}_k'(x)| + \sup_{x\in\Re}|E\hat{f}_k'(x) - f'(x)|.$$

The first term on the right-hand side from (14) is $o_p(1)$. The second term tends to zero by (8), condition b) and Theorem 5 (for the case where $m = 0$) in [6]. Then we have $\lim_{n\to\infty}\sup_{x\in\Re}|\hat{f}_k'(x) - f'(x)| = 0$. Consequently, $\hat{f}_k'(x)$ is uniformly consistent.

The theorem has been proved.

In the following theorem we give orders of the bias and the variance.

Theorem 3 *Assume that a) $f'(x)$ is bounded and continuous, b) there exist functions $H_{2k}(x) > 0$ and $\varepsilon_{2k}(x) > 0$ such that*

$$\left|\Delta_h^{2k} f'(x)\right| \le H_{2k}(x) h^{2k} \quad \text{for} \quad \text{all} \quad |h| \le \varepsilon_{2k}(x) \tag{15}$$

and c) $\int\limits_{-\infty}^{\infty} |K(t)| t^{2k} dt < \infty$. Then, for all $x \in \Re$ and $0 < h_n \le \varepsilon_{2k}(x)$

$$\left|B(\hat{f}'_k(x))\right| \le c h_n^{2k} \left(H_{2k}(x) + \varepsilon_{2k}^{-2k}(x)\right) \tag{16}$$

where the constant c does not depend on x or h_n.

Suppose additionally that d) $\int\limits_{-\infty}^{\infty} |K'(t)|^2 dt < \infty$ and $\int\limits_{-\infty}^{\infty} |M'_k(t)|^2 dt < \infty$ Then, for all $x \in \Re$ and $0 < h_n \le \varepsilon_{2k}(x)$

$$V(\hat{f}'_k(x)) = \frac{1}{n h_n^3} \left(f(x) \int\limits_{-\infty}^{\infty} \left[M'_k(t)\right]^2 dt - h_n [f'(x) + R_{2k}(x, h_n)]^2 \right) \tag{17}$$

where the residual satisfies

$$|R_{2k}(x, h_n)| \le c h_n^{2k} (H_{2k}(x) + \varepsilon_{2k}^{-2k}(x)) \tag{18}$$

with constant c independent of x and h_n.

Proof. Condition c) implies for any $N > 0$

$$\int\limits_{|t|>N} |K(t)|\, dt \le \int\limits_{|t|>N} |K(t)| \left|\frac{t}{N}\right|^{2k} dt \le N^{-2k} \int\limits_{-\infty}^{\infty} |K(t)|\, t^{2k} dt. \tag{19}$$

Using (9) and conditions of Theorem 3, we have

$$\left|B(\hat{f}'_k(x))\right| = \left|\frac{1}{c_{k,0}}\right| \left| \int\limits_{-\infty}^{\infty} K(t) \Delta_{h_n t}^{2k} f'(x) dt \right| \le$$

$$\le c_1 \left(\int\limits_{|h_n t| \le \varepsilon_{2k}(x)} + \int\limits_{|h_n t| > \varepsilon_{2k}(x)} \right) |K(t)| |\Delta_{h_n t}^{2k} f'(x)| dt \le$$

$$\le c_2 \left[H_{2k}(x) \int\limits_{|h_n t| \le \varepsilon_{2k}(x)} |K(t)| (h_n t)^{2k} dt + \sup_{x \in \Re} |f'(x)| \int\limits_{|h_n t| > \varepsilon_{2k}(x)} |K(t)| dt \right].$$

It remains to apply (15) and (19) to obtain (16).

Now, we proceed with derivation of (17). According to (7), we need to evaluate Eu_1^2 and $(Eu_1)^2$. Now,

$$Eu_1 = E(\hat{f}_k'(x)) = f'(x) + B(\hat{f}_k'(x)) = f'(x) + R_{2k}(x, h_n) \tag{20}$$

where $R_{2k}(x, h_n)$ satisfies (18). Now,

$$Eu_1^2 = \left(\frac{1}{h_n^2}\right)^2 \int\limits_{\Re} \left[M_k'\left(\frac{x-t}{h_n}\right)\right]^2 f(t)dt = \frac{1}{h_n^3} \int\limits_{\Re} \left[M_k'(t)\right]^2 f(x - h_n t)dt. \tag{21}$$

Now we show that $\int\limits_{\Re} (M_k'(t))^2 \, dt < \infty$. From (5), we have $M_k'(x) = \sum\limits_{|s|=1}^{k} a_s K'\left(\frac{x}{s}\right)$, where $a_s = -\frac{1}{c_{k,0}} \frac{c_{k,s}}{|s|s}$. Hence, by Hölder's inequality

$$\int\limits_{\Re} (M_k'(x))^2 \, dx = \int\limits_{\Re} \sum_{|s|,|t|=1}^{k} a_s a_t K'\left(\frac{x}{s}\right) K'\left(\frac{x}{t}\right) dx \leq$$

$$\leq \sum_{|s|,|t|=1}^{k} |a_s a_t| \int\limits_{\Re} \left|K'\left(\frac{x}{s}\right)\right| \left|K'\left(\frac{x}{t}\right)\right| dx \leq$$

$$\leq \sum_{|s|,|t|=1}^{k} |a_s a_t| \left(\int\limits_{\Re} |K'|^2 \left(\frac{x}{s}\right) dx\right)^{\frac{1}{2}} \left(\int\limits_{\Re} |K'|^2 \left(\frac{x}{t}\right) dx\right)^{\frac{1}{2}} = c_1 \left(\int\limits_{\Re} |K'(t)|^2 \, dt\right) < \infty$$

because $K' \in L_2(\Re)$.

Note that (17) is a consequence of (20) and (21). In addition, if $f(x) \neq 0$ and for small h_n we can rewrite (17) as

$$V(\hat{f}_k'(x)) = \frac{1}{nh_n^3} \left[f(x) \int\limits_{\Re} (M_k'(t))^2 \, dt + O(h)\right]. \tag{22}$$

The theorem has been proved.

Asymptotically Optimal Bandwidth

In this section we obtain a criterion of the choice of h_n. We consider optimal choice of bandwidth by minimizing the Integrated Mean Squared Error (IMSE),

$$IMSE(\hat{f}'(x)) = \int\limits_{\Re} \left(V(\hat{f}'(x)) + (B(\hat{f}'(x)))^2\right) dx.$$

The value of h_n which minimizes IMSE is called the asymptotically optimal bandwidth. This is done in the following theorem.

Theorem 4 *Let Assumptions of Theorem 3 hold. Suppose that H_{2k}, $\varepsilon_{2k}^{-2k} \in L_2$ and $f, f' \in L_1$, then IMSE is bounded by a function $\varphi(h) = \frac{c_1}{nh^3} + c_2 h^{4k}$. The optimal h_n resulting from the minimization of φ is of order $h_{opt} \asymp n^{-\frac{1}{4k+3}}$.*

Proof. Replacing $V(\hat{f}'(x))$ and $B(\hat{f}'(x))$ in IMSE by their approximations (16) and (22), we get an asymptotic integrated mean squared error, which is denoted by

$$AIMSE = \int\limits_{\Re} \left(\frac{1}{nh_n^3} \left(f(x) \int\limits_{\Re} \left(M_k'(t)\right)^2 dt - h[f'(x) + R_{2k}(x, h_n)]^2 \right) + R_{2k}^2(x, h_n) \right) dx.$$

Under the conditions and minimizing AIMSE over h_n, we have $h_{opt} \asymp n^{-\frac{1}{4k+3}}$.

The theorem has been proved.

References

[1] Emanuel P. On estimation of a Probability Density Function and Mode // Annals of Mathematical Statistics. -1962. -Vol.33. -P. 1065-1076.

[2] Rosenblatt M. Remarks on some non-parametric estimates of a density function // Annals of Mathematical Statistics. -1956. -Vol.27. -P. 832-837.

[3] Bhattacharya P.K. Estimation of a probability density function and its derivatives // Sankhya. -1967. Vol.29. -P. 373-382.

[4] Shuster E.F. Estimation of a probability density function and its derivatives // Annals of Mathematical Statistics. -1969. -Vol.40. -P. 1187-1195.

[5] Silverman B.W. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives // Annals of Mathematical Statistics. -1978. -Vol.6. -P. 177-184.

[6] Mynbaev K.T., Martins-Filho C. Bias reduction in kernel density estimation via Lipschitz condition // Journal of Nonparametric Statistics. -2010. -Vol.22. -P. 219-235.

[7] Jenrich R.I. Asymptotic properties of non-linear least squares estimators // Annals of Mathematical Statistics. -1969. -Vol.40. -P. 633-643.

[8] Tsybakov A.B. Introduction to Nonparametric Estimation // Springer, New York. -2009. -P. 1-76.