Ann. Inst. Statist. Math.
34 (1982), Part A, 479-489

ON ESTIMATION OF A DENSITY AND ITS DERIVATIVES*

K. F. CHENG

(Received May 11, 1981; revised Sept. 1, 1981)

Summary

Let $f_n^{(p)}$ be a recursive kernel estimate of $f^{(p)}$, the pth order derivative of the probability density function f, based on a random sample of size n. In this paper, we provide bounds for the moments of $\|f_n^{(p)}-f^{(p)}\|_{L_2}=\left[\int [f_n^{(p)}(x)-f^{(p)}(x)]^2dx\right]^{1/2}$ and show that the rate of almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_{L_2}$ to zero is $O(n^{-\alpha})$, $\alpha<(r-p)/(2r+1)$, if $f^{(r)}$, $r>p\geqq 0$, is a continuous $L_2(-\infty, \infty)$ function. Similar rate-factor is also obtained for the almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_\infty=\sup_x |f_n^{(p)}(x)-f^{(p)}(x)|$ to zero under different conditions on f.

1. Introduction

Let $(\Omega, \mathcal{A}, P)$ be a probability space on which we observe random variables $X_1, X_2, \cdots, X_n$. Assume that the random variables are independent and identically distributed with common distribution function F and density function f with respect to the Lebesgue measure. For an arbitrary given integer $p\geqq 0$ we in this paper consider a *recursive kernel* estimator $f_n^{(p)}$ of $f^{(p)}$, the pth order derivative of f, based on the random sample. The recursive kernel estimator $f_n^{(p)}$ is given by

$$f_n^{(p)}(x)=n^{-1}\sum_{i=1}^{n} c_i^{-(p+1)}K[(x-X_i)/c_i] ,$$

where K is some suitable kernel function and $\{c_n\}$ is a sequence of nonincreasing positive constants converging to zero as $n\to\infty$. For $p=0$, $f_n(x)$ is a nonparametric estimator of the density function f. This type of estimators was first introduced by Wolverton and Wagner [22] and Yamato [23]. In general, if the convergence of c_n to zero is too

* This work was supported in part by the Research Foundation of SUNY.
AMS 1970 subject classification: Primary 62G05; Secondary 60F15.
Key words and phrases: Recursive kernel, density, derivatives, almost sure convergence, rates.

slow then the estimator will be overly smoothed. On the other hand, if c_n converges to zero too fast, the noise level of the estimator becomes unacceptable. If $c_i=c_n$, $i=1, \cdots, n$, the recursive type estimators become the ordinary kernel type estimators $\hat{f}_n^{(p)}(x)=n^{-1}c_n^{(p+1)}\sum_{i=1}^{n} K[(x-X_i)/c_n]$ of $f^{(p)}(x)$. From the computation point of view it is desirable to use $f_n^{(p)}(x)$, since $f_n^{(p)}$ can be computed recursively.

To study the behavior of the estimator $f_n^{(p)}$, a global measure of deviation of the function $f_n^{(p)}$ from $f^{(p)}$ is given by

$$\|f_n^{(p)}-f^{(p)}\|_\infty=\sup_x |f_n^{(p)}(x)-f^{(p)}(x)| .$$

For $p=0$, the almost sure convergence of $|f_n(x)-f(x)|$ and $\|f_n-f\|_\infty$ to zero were studied by Davies [3] and Deheuvals [5]. A law of the iterated logarithm for $f_n(x)$ was established by Wegman and Davies [21] using the almost sure invariance principle. Sequential procedures for density estimation using f_n and $\hat{f}_n$ were considered by Davies and Wegman [4], Carroll [2], and Wegman and Davies [21]. Regarding the kernel type estimators, the stochastic behavior $\|\hat{f}_n-f\|_\infty$ has been extensively investigated by Parzen [11], Nadaraya [9] and Silverman [12], among others. For $p\geqq 1$, the *almost sure convergence* of $\|\hat{f}_n^{(p)}-f^{(p)}\|_\infty$ to zero has been studied by Singh [13] and Silverman [12]. Similar result is also considered by Walter [18] when the estimator is obtained by using Hermite series method. In this paper, we show that if f is bounded, $f^{(r)}$ is bounded and continuous, $r>p\geqq 0$, and $\mathrm{E}|X|^\delta<\infty$, for some $\delta>0$, then the rate of almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_\infty$ to zero is $o(n^{-(r-p)/(2r+1)}\beta_n \log n)$, where β_n is an arbitrary sequence of positive constants tending to ∞ as $n\to\infty$. It is easy to see that this result can also be generalized to the case when random observations $X_1, X_2, \cdots$, and X_n are $q(n)$-dependent. For $p\geqq 0$, and $c_i=c_n$, $i=1, 2, \cdots, n$, our result for the kernel estimator is better than the similar results developed by Singh [13] and Walter [18] and comparable with the conclusion established in Silverman [12]. However, the generalization of Silverman's result to the recursive kernel type estimators, or $q(n)$-dependent random observations is not obvious.

Along with the distance $\|f_n^{(p)}-f^{(p)}\|_\infty$, another natural and useful measure of the distance between $f_n^{(p)}$ and $f^{(p)}$ is

$$\|f_n^{(p)}-f^{(p)}\|_{L_2}=\left[\int_{-\infty}^{\infty} [f_n^{(p)}(x)-f^{(p)}(x)]^2dx\right]^{1/2} .$$

For $p=0$, attention has been devoted to the rates of convergence of $\mathrm{E}\|\hat{f}_n-f\|_{L_2}^2$ to zero (see review papers by Wegman [19], [20] and Fryer [6]). Recently, for $p\geqq 0$, the exact asymptotic expression of $\mathrm{E}\,(\hat{f}_n^{(p)}(x)$

$-f^{(p)}(x))^2$ has been characterized by Singh [16]. Also, the almost sure convergence of $\|\hat{f}_n-f\|_{L_2}$ to zero has been studied by Nadaraya [10]. However, the almost sure behavior of $\|\hat{f}_n^{(p)}-f^{(p)}\|_{L_2}$ remains unknown for $p\geqq 1$. In the present paper, this question is explored. We show that for general $p\geqq 0$, the rate of almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_{L_2}$ to zero is $O(n^{-\alpha})$, $\alpha<(r-p)/(2r+1)$, if $f^{(r)}$, $r>p\geqq 0$, is a continuous $L_2(-\infty, \infty)$ function. Moreover, the moments $\mathrm{E}\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}$ are characterized. It will be seen that the rates of almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}$ to zero are obtained as consequences of bounds on the moments $\mathrm{E}\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}$.

Estimation of density derivatives arises in *empirical Bayes problems* (see Lin [8]) and also in the problem of estimation of *Fisher information* (see Bhattacharya [1]). Other potential applications of nonparametric estimators of derivatives of a density function can be found in Singh [14].

2. Main results

Following Singh [13], we let $\mathcal{K}(p, r)$ be the class of real valued Borel measurable bounded functions vanishing outside $[a, b]$ (without loss of generality let $a=0$ and $b=1$) such that

$$(2.1)\qquad \left(\frac{1}{j!}\right)\int(-y)^j K(y)dy=\begin{cases}1 & \text{if } j=p\,,\\ 0 & \text{if } j\neq p\,,\quad j=0,1,\cdots,r-1\,,\end{cases}$$

where r is a fixed integer and $r>p$. It is clear that $\mathcal{K}(p, r)$ contains polynomials on $[0, 1]$ satisfying (2.1). For example $K(x)=I\ (0\leqq x\leqq 1)$ $(480x-2700x^2+4320x^3-2100x^4)\in\mathcal{K}(1, 3)$, and satisfies a Lipschitz condition of order 1, where $I(\cdot)$ denotes the indicator function. Other similar examples can be found in Singh [15].

Our first theorem characterizes the rates of strong uniform convergence of the estimator $f_n^{(p)}(x)$.

THEOREM 2.1. *Assume that*

(i) *f is bounded, $f^{(r)}$ is bounded and continuous, $K\in\mathcal{K}(p, r)$, and K satisfies a Lipschitz condition of order* 1;

(ii) *β_n is an arbitrary sequence of positive constants and γ is a positive real number such that $\beta_n\to\infty$ as $n\to\infty$, $n^{\gamma-1}c_n^p\sum_{i=1}^n c_i^{(r-p)}/\beta_n\log n=o(1)$, $n^{\gamma-1}c_n^{-2}/\beta_n\log n=o(1)$, $n^{2\gamma-1}c_n^{-1}=O(1)$ and $n^{\gamma-1}\leqq c_n/4\|K\|_\infty$ for sufficiently large n, where $\|K\|_\infty=\sup_x|K(x)|$;*

(iii) $\mathrm{E}|X|^\delta<\infty$, *for some $\delta>0$. Then*

$$(2.2)\qquad (n^\gamma c_n^p/\beta_n\log n)\|f_n^{(p)}-f^{(p)}\|_\infty\xrightarrow{\text{w.p. 1}}0\,,\qquad \text{as } n\to\infty\,.$$

In particular, let $c_i = d_i i^{-1/(2r+1)}$, *for* $0<\underline{d}\leqq d_i<\bar{d}<\infty$, $i=1, 2, \cdots, n$, *and* $\gamma = r/(2r+1)$, *then*

$$(n^{(r-p)/(2r+1)}/\beta_n \log n)\|f_n^{(p)}-f^{(p)}\|_\infty \xrightarrow{\text{w.p. 1}} 0\,, \qquad \text{as } n\to\infty\,. \tag{2.3}$$

where $\underline{d}$ *and* $\bar{d}$ *are two constants.*

PROOF. The proof of Theorem 2.1 is based on the conclusions developed in Lemmas 2.1 and 2.3. For convenience, we let c denote a generic constant which may not be the same at each appearance.

LEMMA 2.1. *If* $f^{(r)}$ *is bounded and continuous,* $K\in\mathcal{K}(p, r)$ *and* β_n *is an arbitrary sequence of positive constants, then*

$$(n^{\gamma}c_n^p/\beta_n \log n)\|\mathrm{E}\,f_n^{(p)}-f^{(p)}\|_\infty = O\Big(n^{\gamma-1}c_n^p\sum_{i=1}^n c_i^{(r-p)}/\beta_n\log n\Big)\,. \tag{2.4}$$

PROOF. Using Taylor expansion with integral form at the rth term, and the orthogonality properties of K we obtain

$$\|\mathrm{E}\,f_n^{(p)}-f^{(p)}\|_\infty = O\Big(n^{-1}\sum_{i=1}^n c_i^{(r-p)}\Big)\,.$$

Thus (2.4) follows immediately.

LEMMA 2.2. *Suppose that* $f(x)$ *and* $K(x)$ *are bounded functions and* $K\in\mathcal{K}(p, r)$. *If* γ *is a positive real number such that* (i) $n^{2\gamma-1}c_n^{-1}=O(1)$ *and* (ii) $n^{\gamma-1}\leqq c_n/4\|K\|_\infty$ *for sufficiently large* n, *and* β_n *is an arbitrary sequence of positive constants, then for every* $\varepsilon>0$ *and* $x\in R$,

$$\mathrm{P}\,[(n^{\gamma}c_n^p/\beta_n\log n)|f_n^{(p)}(x)-\mathrm{E}\,f_n^{(p)}(x)|>\varepsilon]\leqq cn^{-\varepsilon\beta_n}\,.$$

PROOF. Write

$$\begin{aligned}
&\mathrm{P}\,[(n^{\gamma}c_n^p/\beta_n\log n)(f_n^{(p)}(x)-\mathrm{E}\,f_n^{(p)}(x))>\varepsilon]\\
&\quad=\mathrm{P}\,[n^{\gamma}c_n^p(f_n^{(p)}(x)-\mathrm{E}\,f_n^{(p)}(x))>\varepsilon\beta_n\log n]\\
&\quad\leqq\exp(-\varepsilon\beta_n\log n)\\
&\qquad\times\prod_{i=1}^n \mathrm{E}\,\{\exp[n^{\gamma-1}c_n^p c_i^{-(p+1)}[K[(x-X_i)/c_i]-\mathrm{E}\,K[(x-X_i)/c_i]]]\}\,,
\end{aligned}$$

by using the Chebyschev inequality. For the random variables

$$Z_i = c_i^{-1}\{K[(x-X_i)/c_i]-\mathrm{E}\,K[(x-X_i)/c_i]\}\,,$$

$$|Z_i|\leqq 2c_i^{-1}\|K\|_\infty\,,$$

we may use moment inequality of the exponential form ($\mathrm{E}\exp(\xi(Z-\mathrm{E}\,Z))\leqq\exp(\xi^2\,\mathrm{Var}(Z))$, if $|Z|\leqq\eta$ and $0<\xi<1/(2\eta)$ (see Lamperti [7] pp. 43–44)) in connection with the fact that $\mathrm{Var}(Z_i)\leqq cc_i^{-1}\leqq cc_n^{-1}$ to obtain, for each $i\leqq n$, the following inequality:

$$\begin{aligned}&\mathrm{E}\exp\{n^{r-1}c_n^p c_i^{-(p+1)}[K[(x-X_i)/c_i]-\mathrm{E}\,K[(x-X_i)/c_i]]\}\\&\quad\le\exp(cn^{2(r-1)}c_n^{-1}),\qquad\text{for large } n,\quad i=1,2,\cdots,n.\end{aligned}$$

Thus, for large n,

$$\begin{aligned}\mathrm{P}\,[(n^r c_n^p/\beta_n \log n)(f_n^{(p)}(x)-\mathrm{E}\,f_n^{(p)}(x))>\varepsilon]&\\\le n^{-\varepsilon\beta_n}\exp(c\cdot n^{2r-1}c_n^{-1})&\\\le c\cdot n^{-\varepsilon\beta_n}.&\end{aligned}\tag{2.5}$$

Similarly, we also have

$$\begin{aligned}&\mathrm{P}\,[(n^r c_n^p/\beta_n \log n)(-f_n^{(p)}(x)+\mathrm{E}\,f_n^{(p)}(x))>\varepsilon]\\&\quad\le cn^{-\varepsilon\beta_n},\qquad\text{for large } n.\end{aligned}$$

This in connection with (2.5) establishes the proof of Lemma 2.2.

LEMMA 2.3. *Assume that the conditions of Lemma* 2.2 *are satisfied and K satisfies a Lipschitz condition of order* 1. *If* (i) $\mathrm{E}\,|X|^\delta<\infty$ *for some* $\delta>0$, *and* (ii) $\beta_n\to\infty$ *as* $n\to\infty$ *and* $n^{r-1}c_n^{-2}/\beta_n\log n=o(1)$, *then*

$$(n^r c_n^p/\beta_n \log n)\|f_n^{(p)}-\mathrm{E}\,f_n^{(p)}\|_\infty\xrightarrow{\text{w.p. }1}0,\qquad \text{as } n\to\infty.\tag{2.6}$$

PROOF. Define the set $B_n=\{x\in R\colon |x|\le n^{1/\delta}+1\}$ and consider a set $E_n\subset R$ such that for all $x\in B_n$, there exists $\xi\in E_n$ satisfying $|x-\xi|<1/n$ and E_n contains at most $N_n=2n[n^{1/\delta}+1]+1$ elements. Here $[y]$ denotes the largest integer less than or equal to y. For any $x\in B_n$, we let $y(x)$ be the corresponding element in E_n such that $|x-y(x)|<1/n$. Thus

$$\begin{aligned}&(n^r c_n^p/\beta_n\log n)\sup_{x\in B_n}|f_n^{(p)}(x)-\mathrm{E}\,f_n^{(p)}(x)|\\&\le(n^r c_n^p/\beta_n\log n)\sup_{x\in B_n}|f_n^{(p)}(x)-f_n^{(p)}(y(x))|\\&\quad+(n^r c_n^p/\beta_n\log n)\sup_{x\in B_n}|f_n^{(p)}(y(x))-\mathrm{E}\,f_n^{(p)}(y(x))|\\&\quad+(n^r c_n^p/\beta_n\log n)\sup_{x\in B_n}|\mathrm{E}\,f_n^{(p)}(y(x))-\mathrm{E}\,f_n^{(p)}(x)|\\&=T_{n1}+T_{n2}+T_{n3},\qquad\text{say.}\end{aligned}$$

By utilizing the conclusion of Lemma 2.2,

$$\begin{aligned}\mathrm{P}\,(T_{n2}>\varepsilon)&\le\sum_{y\in E_n}\mathrm{P}\,[(n^r c_n^p/\beta_n\log n)|f_n^{(p)}(y)-\mathrm{E}\,f_n^{(p)}(y)|>\varepsilon]\\&\le c\cdot N_n\cdot n^{-\varepsilon\beta_n},\qquad\text{for large } n.\end{aligned}$$

Since $N_n\cdot n^{-\varepsilon\beta_n}$ is essentially dominated by n^{-2}, thus by Borel-Cantelli lemma we have

$$T_{n2}\xrightarrow{\text{w.p. }1}0\qquad\text{as } n\to\infty.$$

On the other hand, using the Lipschitz property of the function

K and the definition of the set B_n, we obtain, for each $\omega \in \Omega$,

$$T_{n1} \leqq (n^r c_n^p/\beta_n \log n) \sup_{x \in B_n} \left\{ n^{-1} \sum_{i=1}^n c_i^{-(p+1)} |K[(x-X_i)/c_i] - K[(y(x)-X_i)/c_i]| \right\}$$
$$\leqq c \cdot n^{r-1} c_n^{-2}/\beta_n \log n ,$$

which converges to zero as $n \to \infty$. Similarly, it can be shown that $T_{n3} \to 0$ as $n \to \infty$. Consequently,

$$(2.7) \qquad (n^r c_n^p/\beta_n \log n) \sup_{x \in B_n} |f_n^{(p)}(x) - \mathrm{E} f_n^{(p)}(x)| \xrightarrow{\text{w.p. } 1} 0 \qquad \text{as } n \to \infty .$$

Since $\mathrm{E}|X|^\delta < \infty$, for some $\delta > 0$, it can be shown that there exists a set $\Omega_0 \subset \Omega$ such that $p(\Omega_0) = 1$ and for all $\omega \in \Omega_0$ there exists a positive integer N_ω and for all $n \geqq N_\omega$,

$$\max_{1 \leqq i \leqq n} |X_i(\omega)| \leqq n^{1/\delta} .$$

Let N be a positive integer such that for all $n \geqq N$, $c_n \leqq 1$. Consider $x \notin B_n$. If $\omega \in \Omega_0$ then for all $n \geqq \max(N, N_\omega)$, $|x - X_i(\omega)| > 1$, $1 \leqq i \leqq n$, and thus

$$K[(x - X_i(\omega))/c_i] = 0 , \qquad n \geqq i \geqq N .$$

Accordingly,

$$(2.8) \qquad (n^r c_n^p/\beta_n \log n) \sup_{x \notin B_n} |f_n^{(p)}(x)| \xrightarrow{\text{w.p. } 1} 0 \qquad \text{as } n \to \infty .$$

On the other hand, if $x \notin B_n$, then for all $i \geqq N$,

$$K[(x - X_i)/c_i] I(|X_i| \leqq n^{1/\delta}) = 0 ,$$

where $I(\cdot)$ denotes the indicator function. Consequently, for all $n \geqq N$,

$$\sup_{x \notin B_n} \mathrm{E}|f_n^{(p)}(x)| \leqq \sup_{x \notin B_n} n^{-1} \sum_{i=1}^n c_i^{-(p+1)} \{\mathrm{E}|K[(x-X_i)/c_i]| I(|X_i| \leqq n^{1/\delta})$$
$$+ \mathrm{E}|K[(x-X_i)/c_i] I(|X_i| > n^{1/\delta})|\}$$
$$\leqq \sup_{x \notin B_n} \left\{ n^{-1} \sum_{i=1}^N c_i^{-(p+1)} \mathrm{E}|K[(x-X_i)/c_i]| I(|X_i| \leqq n^{1/\delta}) \right\}$$
$$+ n^{-1} \sum_{i=1}^n c_i^{-(p+1)} \|K\|_\infty \mathrm{E}|X|^\delta n^{-1}$$
$$= O(n^{-1} c_n^{-(p+1)}) , \qquad n \to \infty .$$

This shows that

$$(2.9) \qquad (n^r c_n^p/\beta_n \log n) \sup_{x \notin B_n} \mathrm{E}|f_n^{(p)}(x)| \to 0 \qquad \text{as } n \to \infty .$$

Hence (2.6) follows immediately from (2.7), (2.8) and (2.9).

PROOF OF THEOREM 2.1. (2.2) follows easily from the conclusions Lemmas 2.1 and 2.3.

Remarks. (a) Since β_n is an arbitrary but fixed sequence of positive constants tending to ∞ as $n\to\infty$, therefore according to (2.3), the rate of almost sure convergence of $\|f_n^{(p)}-f^{(p)}\|_\infty$ to 0 is close to $o(n^{-(r-p)/(2r+1)}\log n)$.

(b) If A is a bounded subset of R, the set of real numbers, and conditions (i) and (ii) are satisfied, then

$$(n^r c_n^p/\beta_n \log n) \sup_{x\in A} |f_n^{(p)}(x)-f^{(p)}(x)| \xrightarrow{\text{w.p. 1}} 0\,, \qquad \text{as } n\to\infty\,.$$

(c) If $c_n=d_n n^{-1/(2r+1)}$, for $0<\underline{d}\leq d_n\leq\bar{d}<\infty$, and conditions (i) and (iii) are satisfied, then

$$(n^{(r-p)/(2r+1)}/\beta_n \log n)\|\hat{f}_n^{(p)}-f^{(p)}\|_\infty \xrightarrow{\text{w.p. 1}} 0 \qquad \text{as } n\to\infty.$$

In what follows, the almost sure behavior of $\|f_n^{(p)}-f^{(p)}\|_{L_2}$ is established. To do this, we first develop order bounds for the moments of $\|f_n^{(p)}-f^{(p)}\|_{L_2}$.

THEOREM 2.2. *Assume that $K\in\mathcal{K}(p,r)$ and $f^{(r)}$ is a continuous $L_2(-\infty,\infty)$ function. Let s be a positive integer, then*

$$\text{(2.10)}\qquad \mathrm{E}\,\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}=O\Big(\Big(n^{-1}\sum_{i=1}^n c_i^{2(r-p)}\Big)^s\Big)+o(n^{-s}c_n^{-s(2p+1)})\,,$$

and in particular, for $c_i=d_i i^{-1/(2r+1)}$, $0<\underline{d}\leq d_i<\bar{d}<\infty$,

$$\text{(2.11)}\qquad \mathrm{E}\,\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}=O(n^{-2(r-p)s/(2r+1)})\,,$$

where $\underline{d}$ and $\bar{d}$ are two constants.

PROOF. To prove Theorem 2.2 we use the following elementary inequality:

$$\text{(2.12)}\qquad \mathrm{E}\,\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}\leq 2^{2s-1}\{\mathrm{E}\,\|f_n^{(p)}-\mathrm{E}\,f_n^{(p)}\|_{L_2}^{2s}+\|\mathrm{E}\,f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}\}$$

for any positive integer s. The right-most term in (2.12) is simply the sth power of the integrated square bias for which the behavior is developed in Lemma 2.4. The first term on the right-hand side of (2.12) is an sth order analogue of the integrated variance. Lemma 2.5 characterizes the behavior of this term.

LEMMA 2.4. *Assume that $K\in\mathcal{K}(p,r)$ and $f^{(r)}$ is a continuous $L_2(-\infty,\infty)$ function. Let s be a positive integer. Then*

$$\text{(2.13)}\qquad \|\mathrm{E}\,f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}=O\Big(\Big(n^{-1}\sum_{i=1}^n c_i^{2(r-p)}\Big)^s\Big)\,.$$

PROOF. Using Taylor expansion with integral form of the remainder at the rth term and the orthogonality properties of K, we have

$$\mathrm{E}f_n^{(p)}(x)$$
$$=f^{(p)}(x)+((r-1)!)^{-1}n^{-1}\sum_{i=1}^{n}c_i^{-p}\int_0^1(-c_iy)^rK(y)\int_0^1(1-t)^{r-1}f^{(r)}(x-c_iyt)dt\,dy\,.$$

Thus by virtue of Holder's inequality and Fubini's theorem we obtain

$$\begin{aligned}
&\|\mathrm{E}f_n^{(p)}-f^{(p)}\|_{L_2}^2\\
&\leqq((r-1)!)^{-2}n^{-1}\\
&\quad\times\sum_{i=1}^{n}c_i^{-2p}\int_{-\infty}^{\infty}\left\{\int_0^1(-c_iy)^rK(y)\int_0^1(1-t)^{r-1}f^{(r)}(x-c_iyt)dt\,dy\right\}^2dx\\
&\leqq((r-1)!)^{-2}n^{-1}\sum_{i=1}^{n}c_i^{2(r-p)}\int_0^1\int_0^1K^*y^{2r}(1-t)^{2r-2}\|f^{(r)}\|_{L_2}^2|K(y)|dt\,dy\\
&=((r-1)!)^{-2}n^{-1}\sum_{i=1}^{n}c_i^{2(r-p)}K^*\|f^{(r)}\|_{L_2}^2\int_0^1y^{2r}|K(y)|dy\int_0^1(1-t)^{2r-2}dt\\
&=O\left(n^{-1}\sum_{i=1}^{n}c_i^{2(r-p)}\right),\qquad n\to\infty\,,
\end{aligned}$$

where $K^*=\int_0^1|K(y)|dy$ and $\|f^{(r)}\|_{L_2}^2=\int_{-\infty}^{\infty}(f^{(r)}(x))^2dx$. This completes the proof of the lemma.

LEMMA 2.5. *Assume $K\in\mathcal{K}(p,r)$ and let s be a positive integer. Then*

$$\mathrm{E}\|f_n^{(p)}-\mathrm{E}f_n^{(p)}\|_{L_2}^{2s}=O(n^{-s}\cdot c_n^{-s(2p+1)})\,. \tag{2.14}$$

PROOF. Define $Y_i(x)=c_i^{-(p+1)}\{K[(x-X_i)/c_i]-\mathrm{E}\,K[(x-X_i)/c_i]\}$. Then

$$\int_{-\infty}^{\infty}(f_n^{(p)}(x)-\mathrm{E}f_n^{(p)}(x))^2dx=n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty}(Y_i(x)Y_j(x))dx\,.$$

Here for each i and j,

$$\begin{aligned}\int_{-\infty}^{\infty}|Y_{ni}(x)Y_{nj}(x)|dx&\leqq4K^*\|K\|_\infty c_j^{-p}c_i^{-(p+1)}\\&\leqq4K^*\|K\|_\infty c_n^{-(2p+1)}\,.\end{aligned} \tag{2.15}$$

Furthermore,

$$\begin{aligned}&\mathrm{E}\|f_n^{(p)}-\mathrm{E}f_n^{(p)}\|_{L_2}^{2s}\\&=n^{-2s}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\cdots\sum_{i_s=1}^{n}\sum_{j_s=1}^{n}\mathrm{E}\left\{\prod_{k=1}^{s}\int_{-\infty}^{\infty}Y_{i_k}(x_k)Y_{j_k}(x_k)dx_k\right\}\end{aligned} \tag{2.16}$$

and

$$\mathrm{E}\left\{\prod_{k=1}^{s}\int_{-\infty}^{\infty} Y_{i_k}(x_k)Y_{j_k}(x_k)dx_k\right\}=\int\cdots\int \mathrm{E}\left\{\prod_{k=1}^{s} Y_{i_k}(x_k)Y_{j_k}(x_k)\right\}dx_1\cdots dx_k\,,$$

by using the fact of (2.15) and Fubini's theorem. By independence of $Y_i(x)$'s, $1\leqq i\leqq n$, the expectation in the integrand is zero except in the case that each index in the list $i_1, j_1, \cdots, i_s, j_s$ appears at least twice. In this case, the number of distinct elements in the set $\{i_1, j_1, \cdots, i_s, j_s\}$ is $\leqq s$. It follows that the number of ways to choose $i_1, j_1, \cdots, i_s, j_s$ such that the expectation in (2.16) is nonzero is $O(n^s)$. Moreover, these nonzero expectations are uniformly $O(c_n^{-(2p+1)s})$. Hence

$$\mathrm{E}\,\|f_n^{(p)}-\mathrm{E}\,f_n^{(p)}\|_{L_2}^{2s}=O(n^{-s}c_n^{-(2p+1)s})\,.$$

PROOF OF THEOREM 2.2. The proof of (2.10) follows easily from the conclusions of Lemmas 2.4 and 2.5 in conjunction with the relation (2.12).

Remarks. (a) If $c_n=d_n n^{-1/(2r+1)}$ and $0<\underline{d}<d_n\leqq\bar{d}<\infty$, then

$$\mathrm{E}\,\|\hat{f}_n^{(p)}-f^{(p)}\|_{L_2}^{2s}=O(n^{-2(r-p)s/(2r+1)})\,.$$

(b) For $s=1$, the rate of convergence of the mean integrated square error becomes $n^{-2(r-p)/(2r+1)}$, if $c_i=d_i i^{-1/(2r+1)}$. Walter [17] shows that $\mathrm{E}\,\|\tilde{f}_n^{(p)}-f^{(p)}\|_{L_2}^2=O(n^{(p/r)+(5/6r)-1})$, if $\tilde{f}_n^{(p)}$ is an estimator based on Hermite series method, and r is some positive integer such that $(x-D)^r f \in L_2(-\infty,\infty)$, and $0\leqq p<r$. Clearly, our rate-factor is better than $n^{(p/r)+(5/6r)-1}$.

By virtue of the conclusion of Theorem 2.2, it is easy to verify

THEOREM 2.3. *Assume that $K\in\mathcal{K}(p, r)$, $f^{(r)}$ is a continuous $L_2(-\infty, \infty)$ function, and $c_i=d_i i^{-1/(2r+1)}$, where $0<\underline{d}\leqq d_i\leqq\bar{d}<\infty$, $i=1, 2, \cdots, n$. Then for $\alpha<(r-p)/(2r+1)$,*

$$n^{\alpha}\|f_n^{(p)}-f^{(p)}\|_{L_2}\xrightarrow{\text{w.p. 1}}0\,,\qquad as\ n\to\infty\,. \tag{2.17}$$

PROOF. Applying the Chebyschev inequality and Theorem 2.2, we have, for any $\varepsilon>0$,

$$\begin{aligned}\mathrm{P}\,(n^{\alpha}\|f_n^{(p)}-f^{(p)}\|_{L_2}>\varepsilon)&\leqq\varepsilon^{-2s}n^{2\alpha s}\,\mathrm{E}\,\|f_n^{(p)}-f^{(p)}\|_{L_2}^{2s}\\&=O(n^{2s(\alpha-(r-p))/(2r+1)})\,,\end{aligned}$$

where s is any positive integer. Since $\alpha<(r-p)/(2r+1)$, thus (2.17) easily follows by the Borel-Cantelli lemma and the fact that s may be chosen arbitrarily large.

Remark. It is obvious that by utilizing the same argument we also have

$$n^{\alpha}\|\hat{f}_n^{(p)}-f^{(p)}\|_{L_2}\xrightarrow{\text{w.p. }1}0, \qquad \text{as } n\to\infty\,,$$

for $\alpha<(r-p)/(2r+1)$, if $c_n=d_n n^{-1/(2r+1)}$.

Acknowledgement

The advice of the referee has been of great assistance in the preparation of this paper.

STATE UNIVERSITY OF NEW YORK AT BUFFALO

REFERENCES

[1] Bhattacharya, P. K. (1967). Estimation of a probability density function and its derivatives, *Sankhyā*, A, **29**, 373-382.

[2] Carroll, R. J. (1976). On sequential density estimation, *Z. Wahrscheinlichkeitsth.*, **36**, 137-151.

[3] Davies, H. I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist.*, **15**, 49-54.

[4] Davies, H. I. and Wegman, E. J. (1975). Sequential nonparametric density estimation, *IEEE Trans. Inf. Theory*, IT-**21**, 619-628.

[5] Deheuvals, P. (1974). Conditions necessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C.R. Acad. Sci. Paris*, A, **278**, 1217-1220.

[6] Fryer, M. J. (1977). A review of some non-parametric methods of density estimation, *J. Inst. Math. Appl.*, **20**, 335-354.

[7] Lamperti, J. (1966). *Probability*, W. A. Benjamin, Inc. N.Y.

[8] Lin, P. E. (1975). Rates of convergence in empirical Bayes problems: Continuous case, *Ann. Statist.*, **3**, 155-164.

[9] Nadaraya, E. A. (1965). On nonparametric estimates of density functions and regression curves, *Theory Prob. Appl.*, **10**, 186-190.

[10] Nadaraya, E. A. (1973). On convergence in the L_2-norm of probability density estimates, *Theory Prob. Appl.*, **18**, 808-811.

[11] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.

[12] Silverman, B. S. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Ann. Statist.*, **6**, 177-184.

[13] Singh, R. S. (1977a). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density, *Ann. Statist.*, **5**, 394-400.

[14] Singh, R. S. (1977b). Applications of estimators of a density and its derivatives to certain statistical problems, *J.R. Statist. Soc.*, B, **39**, 357-363.

[15] Singh, R. S. (1979). On necessary and sufficient conditions for uniform strong consistency of estimators of a density and its derivatives, *J. Multivariate Anal.*, **9**, 157-164.

[16] Singh, R. S. (1981). On the exact asymptotic behavior of estimators of a density and its derivatives, *Ann. Statist.*, **9**, 453-456.

[17] Walter, G. G. (1977). Properties of Hermite series estimation of probability density, *Ann. Statist.*, **5**, 1258-1264.

[18] Walter, G. G. (1980). Addendum to "Properties of Hermite series estimation of probability density", *Ann. Statist.*, **8**, 454-455.

[19] Wegman, E. J. (1972a). Nonparametric probability density estimation: I. A sum-

mary of available methods, *Technometrics*, **14**, 533-546.
[20] Wegman, E. J. (1972b). Nonparametric probability density estimation: II. A comparison of density estimation methods, *J. Statist. Comp. Simul.*, **1**, 225-245.
[21] Wegman, E. J. and Davies, H. I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Statist.*, **7**, 316-327.
[22] Wolverton, C. T. and Wagner, T. J. (1969). Asymptotically optimal discriminant functions for pattern classification, *IEEE Trans. Inf. Theory*, IT-**15**, 258-265.
[23] Yamato, H. (1971). Sequential estimation of a continuous probability density and mode, *Bull. Math. Statist.*, **14**, 1-12.