The Annals of Statistics
1977, Vol. 5, No. 2, 394-399

IMPROVEMENT ON SOME KNOWN NONPARAMETRIC UNIFORMLY CONSISTENT ESTIMATORS OF DERIVATIVES OF A DENSITY

BY R. S. SINGH[1]

Indian Statistical Institute, New Delhi

Based on a random sample from a univariate distribution with density f, this note exhibits a class of kernel estimators of the pth order derivative $f^{(p)}$ of f, $p \geqq 0$ fixed. These estimators improve some known estimators of $f^{(p)}$ by weakening the conditions, sharpening the rates of convergence, or both for the properties of strong consistency, asymptotic unbiasedness and mean square consistency, each uniform on the real line.

1. Introduction. In recent years estimation of a Lebesgue density f by the kernel method (apparently introduced by Rosenblatt (1956)) has become common in the literature (e.g., Parzen (1962), Nadaraya (1965), Bhattacharya (1967), Schuster (1969) and Singh (1974)). It is also known that a kernel estimator $\tilde{f}$ of f is (under certain conditions on the kernel and the scaling parameter involved) asymptotically unbiased, mean square consistent, strongly consistent and asymptotically normally distributed at every continuity point of f. Presumably tempted by these properties of $\tilde{f}$, Bhattacharya (1967) (and later Schuster (1969)) suggested estimation of $f^{(p)}$, the pth order derivative of f, by $\tilde{f}^{(p)}$, the pth order derivative of $\tilde{f}$. Intuitively one might think that this is the best way of estimating $f^{(p)}$ since $\tilde{f}$ as an estimator of f has the abovementioned desirable properties. This paper, however, suggests otherwise by exhibiting kernel estimators $\hat{f}^{(p)}$ of $f^{(p)}$ which are not necessarily pth order derivatives of $\hat{f} = \hat{f}^{(0)}$, and yet $\hat{f}^{(p)}$ possess some desirable asymptotic properties *under* conditions weaker than those imposed by Bhattacharya (1967) and Schuster (1969) for similar properties.

Another widely used method of nonparametric estimation of a density is the orthogonal series method. A notable work on this is due to Schwartz (1967). The present note shows that compared to the mean squared error (MSE) of the orthogonal series estimator f^* (of f) of Schwartz, the MSE of our kernel estimator $\hat{f}$ (of f) converges *under* much weaker conditions, to zero at a considerably faster rate.

Throughout this paper, $X_1, \cdots, X_n$ are independent real valued random variables with Lebesgue density f. *For a real valued function t on the real line,* $\sup_x |t(x)|$ *is denoted by* $||t||$.

2. Known results. The most general form (Rosenblatt (1956), Parzen (1962),

Received March 1976; revised July 1976.

[1] Now at the University of Guelph.

AMS 1970 *subject classification.* Primary 62G05.

Key words and phrases. Nonparametric estimation, pth order derivative of a density, uniform convergence, rates.

Nadarya (1965), Bhattacharya (1967), Schuster (1969) and Singh (1974)) of a kernel estimator $\tilde{f}$ of f is given by

$$\tilde{f}(x) = (nh)^{-1} \textstyle\sum_1^n K_0\left(\frac{x - X_j}{h}\right), \tag{2.0}$$

where $0 < h = h(n)$ is the scaling parameter converging to 0 as $n \to \infty$ and K_0 is a real valued function on the real line. Bhattacharya (1967) (and later Schuster (1969) estimated $f^{(p)}$, the pth order derivative of f, by $\tilde{f}^{(p)}$, the pth order derivative of $\tilde{f}$, i.e., by

$$\tilde{f}^{(p)}(x) = (nh^{p+1})^{-1} \textstyle\sum_1^n K_0^{(p)}\left(\frac{x - X_j}{h}\right), \tag{2.1}$$

where $K_0^{(p)}$ is the pth order derivative of K_0.

THEOREM 2.1 (Bhattacharya (1967)). *If $f, f^{(1)}, \cdots, f^{(p+1)}$ all are bounded, then under certain conditions on K_0* (see [1], page 374)

$$\|E\tilde{f}^{(p)} - f^{(p)}\| = O(h)\,. \tag{2.2}$$

THEOREM 2.2 (Schuster (1969)). *If $f, f^{(1)}, \cdots, f^{(p+1)}$ all are bounded and $h = n^{-1/(2p+4)}$, then under certain conditions on K_0* (see [7], page 1188) *for every $\varepsilon > 0$*

$$\|\tilde{f}^{(p)} - f^{(p)}\| = o(n^{-(1+\varepsilon)/(2p+4)}) \quad \text{w.p. } 1\,. \tag{2.3}$$

REMARK 2.2.1. In fact, if Schuster takes $h = (n^{-1} \log\log n)^{1/(2p+4)}$ and in the proof of his theorem makes use of Theorem 2 of Kiefer (1961), he would get lhs of (2.3) $= O(n^{-1} \log\log n)^{1/(2p+4)}$ w.p. 1, a slight improvement in the rate of convergence.

THEOREM 2.3 (Schwartz (1967)). *Let f be of bounded variation and for an integer $r \geqq 3$, $f^{(r)}$ exist and for each $j = 0, 1, \cdots, r$, $\int (x^j f^{(r-j)}(x))^2\, dx < \infty$. Let f^* be defined as in* (3.1) *of* [6], (f^* depends on $X_1, \cdots, X_n$, n and r). *Then*

$$\|E(f^* - f)^2\| = O(n^{-(r-2)/r})\,. \tag{2.4}$$

Our kernel estimators (to be introduced in the following section) of $f^{(p)}$ possess properties (2.2) and (2.3) even if *only* $f^{(p+1)}$ is bounded. In contrast to Theorem 2.3, if *only* f is bounded and for an integer $r > 0$ $\int (f^{(r)})^2 < \infty$, then the lhs of (2.4) for our kernel estimators of f is $O(n^{-(2r-1)/(2r)})$. *Proofs are extremely simple.*

3. Kernel estimators of $f^{(p)}$, results and rates of convergence. Let $r > p$ be a fixed integer. Let $\mathscr{K}$ be the class of all real valued Borel measurable bounded functions K vanishing outside of (0, 1) such that

$$\begin{aligned} \frac{1}{j!} \int y^j K(y)\, dy &= 1 \quad \text{if } j = p \\ &= 0 \quad \text{if } j \neq p\,,\ j = 0, 1, \cdots, r-1\,. \end{aligned} \tag{3.0}$$

($\mathscr{K}$ contains polynomials on (0, 1) satisfying (3.0).) Let $0 < h = h(n) < 1$ converge to zero as $n \to \infty$. For a fixed K in $\mathscr{K}$ definite

$$\hat{f}^{(p)}(x) = (nh^{p+1})^{-1} \sum_1^n K\left(\frac{X_j - x}{h}\right). \tag{3.1}$$

It may be noted that the kernel introduced in (3.0) could be negative, thus leading to a negative estimate for the density. But this is the price one pays for faster rates of convergence for kernel estimates of a density.

THEOREM 3.1. *If* $f^{(r)}$ *is bounded, then*

$$||E\hat{f}^{(p)} - f^{(p)}|| = O(h^{r-p}). \tag{3.2}$$

And if for a $t > 1$, $\int |f^{(r)}|^t < \infty$, *then*

$$\text{lhs } of \text{ (3.2)} = O(h^{r-p-t^{-1}}). \tag{3.3}$$

PROOF. Since X_j are i.i.d. with density f, by a use of the transformation theorem

$$Ef^{(p)}(x) = h^{-p} \int K(y) f(x + hy)\, dy. \tag{3.4}$$

Since $f^{(r)}$ exists and is integrable on $[x, x + h)$, expanding $f(x + hy)$ at x in hy with integral form of the remainder at the rth term and then making use of the orthogonality properties (3.0) of K, we get

$$E\hat{f}^{(p)}(x) = f^{(p)}(x) + ((r-1)!\, h^p)^{-1} \int K(y)\{\textstyle\int_x^{x+hy} (x+hy-t)^{r-1} f^{(r)}(t)\, dt\}\, dy. \tag{3.5}$$

Since K vanishes off (0, 1) and is bounded (say) by M, we get from (3.5)

$$|E\hat{f}^{(p)}(x) - f^{(p)}(x)| \leqq Mh^{r-p-1} \textstyle\int_x^{x+h} |f^{(r)}|. \tag{3.6}$$

Now (3.2) and (3.3) follow from (3.6), since $\int_x^{x+h} |f^{(r)}| \leqq h||f^{(r)}||$, and by Hölder's inequality for $t > 1$, $\int_x^{x+h} |f^{(r)}| \leqq \{h^{(t-1)} \int |f^{(r)}|^t\}^{1/t}$. □

The result of Theorem 2.1 (for which boundedness of all $f, f^{(1)}, \cdots, f^{(p+1)}$ is assumed) is similar to our result (3.2) with $r = p + 1$ (for which the boundedness of only $f^{(r)}$ is assumed).

THEOREM 3.2. *Let* K *in* (3.1) *be continuous and of bounded variation. If* (3.2) *holds, then taking* $h = n^{-1/(2r+2)}$,

$$||\hat{f}^{(p)} - f^{(p)}|| = O(n^{-(r-p)/(r+1)} \log\log n)^{\frac{1}{2}} \quad \text{w.p. } 1. \tag{3.7}$$

And if (3.3) *holds, then taking* $h = n^{-w/2}$ *with* $w = (r + 1 - t^{-1})^{-1}$,

$$\text{lhs } of \text{ (3.7)} = O(n^{-w(r-p-t^{-1})} \log\log n)^{\frac{1}{2}} \quad \text{w.p. } 1. \tag{3.8}$$

NOTE. Since $\hat{f}^{(p)}$ depend on h, $\hat{f}^{(p)}$ in (3.7) are different from those in (3.8).

PROOF. Denote $K\{(\cdot - x)/h\}$ by $Z_x(\cdot)$. Let F^* be the empirical distribution function of $X_1, \cdots, X_n$. Let $F = E(F^*)$. Then, since $h^{p+1}\hat{f}^{(p)}(x) = \int Z_x\, dF^*$, we have

$$h^{p+1}(\hat{f}^{(p)}(x) E\hat{f}^{(p)}(x)) = \int Z_x\, d(F^* - F) = \int (F - F^*)\, dZ_x \tag{3.9}$$

where the second equation follows by integration by parts together with $K(t) \equiv 0$ for $t \notin (0, 1)$. Since K is of bounded variation with total variation of (say) k, $\int |dZ_x| \leqq k < \infty$, and we have from (3.9)

$$h^{p+1}\|\hat{f}^{(p)} - E\hat{f}^{(p)}\| \leqq k\|F^* - F\| = O(n^{-1} \log\log n)^{\frac{1}{2}} \quad \text{w.p. } 1 \tag{3.10}$$

where the equation follows from Theorem 2 of Kiefer (1961).

Since $(\hat{f}^{(p)} - f^{(p)}) = (\hat{f}^{(p)} - E\hat{f}^{(p)}) + (E\hat{f}^{(p)} - f^{(p)})$, (3.7) and (3.8) follow from (3.10) and their respective assumptions. □

The result of Theorem 2.2 (which assumes the boundedness of all $f, f^{(1)}, \cdots, f^{(p+1)}$) is obtained from (3.7) with $r = p + 1$ (for which the boundedness of *only* $f^{(p+1)}$ is assumed).

To *improve* rates in (3.7) and (3.8) we have

REMARK 3.2.1. From (3.2) and (3.10), if $h^2 = (n^{-1} \log\log n)^{1/(r+1)}$, then (3.7) is strengthened to

$$\text{lhs of (3.7)} = O(n^{-1} \log\log n)^{(r-p)/(2r+2)} \quad \text{w.p. } 1\,. \tag{3.7$'$}$$

And from (3.3) and (3.10), if $h^2 = (n^{-1} \log\log n)^w$, with $w = (r + 1 - t^{-1})^{-1}$, then (3.8) is strengthened to

$$\text{lhs of (3.7)} = O(n^{-1} \log\log n)^{w(r-p-t^{-1})/2} \quad \text{w.p. } 1\,. \tag{3.8$'$}$$

It may be noted that rates in (3.8) and (3.8)′ increase with t.

THEOREM 3.3. *If* (3.6) *holds, then*

$$E(\hat{f}^{(p)}(x) - f^{(p)}(x))^2 \leqq M^2(h^{r-p-1} \textstyle\int_x^{x+h} |f^{(r)}|)^2 + (nh^{2p+2})^{-1} \int_x^{x+h} f \tag{3.11}$$

where M is a bound for $|K|$ in (3.1).

PROOF. Since X_j are i.i.d. with density f,

$$\begin{aligned}(nh^{2p+2}) \text{ Variance } (\hat{f}^{(p)}(x)) &= \text{Variance}\left(K\left(\frac{X_1 - x}{h}\right)\right) \\ &\leqq \textstyle\int K^2((y - x)/h)f(y)\, dy \\ &\leqq M^2 \textstyle\int_x^{x+h} f(y)\, dy\end{aligned} \tag{3.12}$$

since $|K| \leqq M$ and vanishes off (0, 1). Note that lhs of (3.11) = (lhs (3.6))^2 + Var $(\hat{f}^{(p)}(x))$. Thus (3.11) follows from (3.6) and (3.12). □

The following corollaries are to Theorem 3.3, each of which gives rates for $\|E(\hat{f}^{(p)} - f^{(p)})^2\|$.

COROLLARY 3.3.1. *If for a* $t \geqq 1$ *and a* $t_* \geqq 1$, $\int |f^{(r)}|^t < \infty$ *and* $\infty > \int |f|^{t_*}$ (=1 if $t_* = 1$), *then with* $h = n^{-s}$, *where* $s = (2r - 2t^{-1} + 1 + t_*^{-1})^{-1}$,

$$\|E(\hat{f}^{(p)} - f^{(p)})^2\| = O(n^{-2s(r-p-t^{-1})})\,. \tag{3.13}$$

PROOF. By Hölder inequality $\int_x^{x+h} |f^{(r)}| \leqq (h^{(t-1)} \int |f^{(r)}|^t)^{1/t}$ for $t \geqq 1$, and $\int_x^{x+h} f \leqq (h^{t_*-1} \int (f)^{t_*})^{1/t_*}$ for $t_* \geqq 1$. Now (3.11) together with these inequalities gives (3.13). □

Note that the larger the t and t_* in the above corollary, the faster the rate in (3.13). This rate is improved by each of the next three corollaries.

COROLLARY 3.3.2. *If for a* $t \geqq 1$, $\int |f^{(r)}|^t < \infty$ *and* f *is bounded, then with* $h = n^{-v}$ *where* $v = (2r + 1 - 2t^{-1})^{-1}$,

$$\text{lhs of } (3.13) = O(n^{-2v(r-p-t^{-1})}) . \tag{3.14}$$

PROOF. Since $h^{-1} \int_x^{x+h} f$ is bounded in x, the proof of (3.14) follows from that of (3.13). □

COROLLARY 3.3.3. *If* $f^{(r)}$ *is bounded and for a* $t_* \geqq 1$, $\infty > \int (f)^{t_*}$ $(=1$ *with* $t_* = 1)$, *then with* $h = n^{-u}$ *where* $u = (2r + 1 + t_*^{-1})^{-1}$,

$$\text{lhs of } (3.13) = O(n^{-2u(r-p)}) . \tag{3.15}$$

PROOF. Since $h^{-1} \int_x^{x+h} |f^{(r)}|$ is bounded in x, the proof of (3.15) follows from that of (3.13). □

The rates in (3.14) and (3.15) increase with t and t_*. Rates in (3.15) are better than those in (3.14) for $t = t_* = 1$. The same conclusion holds for $t = 2$ and $t_* = 1$ iff $r < 2p + 1$, for $t = t_* = 2$ iff $r \leqq 3p + 1$; and for $t = 1$ and $t_* = 2$ iff $r \leqq 5p + 3$.

The rate obtained in (3.16) below is the *sharpest* among all those obtained in (3.13), (3.14) and (3.15). Rates in (3.13), (3.14) and (3.15) approach to that in (3.16) as t, t_* there approach to infinity.

COROLLARY 3.3.4. *If* f *and* $f^{(r)}$ *are bounded, then with* $h = n^{-1/(1+2r)}$,

$$\text{lhs of } (3.13) = O(n^{-2(r-p)/(1+2r)}) . \tag{3.16}$$

PROOF. The proof follows from (3.11) since $\int_x^{x+h} |f^{(r)}| \leqq h||f^{(r)}||$ and $\int_x^{x+h} f \leqq h||f||$. □

Note that the $\hat{f}^{(p)}$ in (3.13), (3.15) and (3.16) are not the same since the h there are different and $\hat{f}^{(p)}$ depends on h.

The rate obtained in Theorem 2.3 (for which it is assumed that f is of bounded variation (and hence necessarily bounded) and for an integer $r \geqq 3$, $f^{(r)}$ exists and for each $j = 0, 1, \cdots, r$, $\int (x^j f^{(r-j)}(x))^2\, dx < \infty$) is improved *significantly* (especially when higher derivatives of f do not exist) by Corollary 3.3.2 with $p = 0$ and $t = 2$ (for which the *only* assumptions are that f is bounded and for an integer $r > 0$, $\int |f^{(r)}|^2 < \infty$).

4. Some final remarks. Those interested in further properties (such as asymptotic normality or integrated mean square consistency) of estimators (3.1) may look at Chapter 1, Singh (1974). A generalization of the work reported here to the independent nonidentically distributed case is considered in Singh (1975). Kernel estimation of mixed partial derivatives of a multivariate density is treated in Singh (1976), but unlike here, the author fails to obtain any explicit rate of convergence for asymptotic unbiasedness or mean square consistency of the estimators there. The treatment in the latter two, however, cannot be expected to be simpler than that in the present one.

Estimators of f or $f^{(1)}$ are widely used in solving a number of statistical problems, e.g., see Bhattacharya (1967), Nadaraya (1965) and Singh (1974, Chapter 2). Some important applications of estimators of $f^{(p)}$ for p larger than 1 can be found in Singh and Tracy (1975).

Acknowledgment. The author wishes to thank Professor James Hannan for his various suggestions and comments on the original results.

REFERENCES

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

[2] KIEFER, J. (1961). On large derivations of the empirical distribution of vector chance variables and law of interation logarithm. *Pacific J. Math.* **11** 649-660.

[3] NADARAYA, E. A. (1965). On nonparametric estimates of density function and regression curves. *Theor. Probability Appl.* **10** 186-190.

[4] PARZON, EMANUEL (1962). On the estimation of probability density and mode. *Ann. Math. Statist.* **33** 1065-1076.

[5] ROSENBLATT, MURRAY (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832-837.

[6] SCHWARTZ, STUART C. (1967). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* **38** 1261-1265.

[7] SCHUSTER, EUGENE F. (1969) Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187-1195.

[8] SINGH, R. S. (1974). Estimation of derivatives of average of densities and sequence-compound estimation in exponential families. RM-18, Department of Statistics and Probability, Michigan State Univ.

[9] SINGH, R. S. (1975). Nonparametric estimation of derivatives of average of densities with convergence rates and applications. Unpublished.

[10] SINGH, R. S. and TRACY, D. S. (1975). Strongly consistent estimators of kth order regression curves and rates of convergence. Unpublished.

[11] SINGH, R. S. (1976). Nonparametric estimation of mixed partial derivatives of a multivariate density. *J. Multivariate Analysis* **6** 111-122.

DEPARTMENT OF MATHEMATIC STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO CANADA