The Annals of Statistics
1981, Vol. 9, No. 2, 453–456

ON THE EXACT ASYMPTOTIC BEHAVIOR OF ESTIMATORS OF A DENSITY AND ITS DERIVATIVES[1]

BY R. S. SINGH

University of Guelph

For an integer $p \geq 0$, Singh has proposed a class of kernel estimators $\hat{f}^{(p)}$ of the pth order derivative $f^{(p)}$ of a density f. This paper examines the detailed asymptotic behavior of these estimators. In particular, asymptotically equivalent expressions for the bias $(E\hat{f}^{(p)} - f^{(p)})$, the mean squared error $E(\hat{f}^{(p)} - f^{(p)})^2$ and the error $(\hat{f}^{(p)} - f^{(p)})$ are obtained, which in turn give exact rates of convergence of these terms to zero.

1. Introduction and preliminaries. Let $\mathbf{X}_n = (X_1, \cdots, X_n)$ be a random sample from a univariate population with unknown probability density function (pdf) f, and let $p \geq 0$ be an arbitrary but fixed integer. Based on $\mathbf{X}_n$, Singh (1977) exhibited a class of kernel estimators $\hat{f}^{(p)}$ of $f^{(p)}$, the pth order derivative of f, and proved various asymptotic properties of $\hat{f}^{(p)}$ concerning their bias, error and mean squared error. Rates of convergence for each of these terms, uniform on the real line, are also obtained there, bringing improvements over the corresponding results of Bhattacharya (1967), Schwartz (1967) and Schuster (1969). This note is formulated to study the exact asymptotic behavior of these improved estimators of Singh (1977). More precisely, we will obtain here the exact asymptotic expressions (and hence also the exact rates of convergence) for the bias $(E\hat{f}^{(p)} - f^{(p)})$, the mean squared error $E(\hat{f}^{(p)} - f^{(p)})^2$ and the error $(\hat{f}^{(p)} - f^{(p)})$ of estimators $\hat{f}^{(p)}$ of $f^{(p)}$. Results on $\text{Cov}(\hat{f}^{(p)}(x_1), \hat{f}^{(p)}(x_2))\, x_1 \neq x_2$, are also obtained.

We will briefly reintroduce estimators $\hat{f}^{(p)}$. Let $r > p$ be an integer, and K be a real valued bounded function vanishing off (0, 1) such that

$$\frac{1}{j!}\int y^j K(y)\,dy = 1 \quad \text{if} \quad j = p \qquad (1.1)$$
$$= 0 \quad \text{if} \quad j \neq p, j = 0, 1, \ldots, r-1.$$

Let $0 < h_n = h \downarrow 0$ as $n \to \infty$. (The restriction $h < 1$ put in Singh (1977) is in fact unnecessary for the results obtained there.) Let

$$Y_j(x) = h^{-p-1} K\left(\frac{X_j - x}{h}\right). \qquad (1.2)$$

Then Singh's (1977) estimator of $f^{(p)}$ at x is

$$\hat{f}^{(p)}(x) = n^{-1} \textstyle\sum_{j=1}^{n} Y_j(x). \qquad (1.3)$$

Silverman (1978) has considered estimators $\hat{f}^{(0)}$ (with different kernels) of f and Bhattacharya's (1967) estimators $\partial^p \hat{f}^{(0)}(x)/\partial x^p$ of $f^{(p)}(x)$. His objective there has been to prove uniform strong consistency of these estimators under minimal conditions on h rather than to investigate the exact asymptotic behavior of the terms like those mentioned earlier concerning estimators of $f^{(p)}$. Silverman has obtained the minimal conditions on h for the purpose with some stringent restrictions on the kernel; see Singh (1979a). How estimators $\hat{f}^{(p)}$ improve Bhattacharya's (1967) estimators of $f^{(p)}$ is discussed in Singh (1977).

As mentioned in Singh (1977), kernels (1.1) could be negative, thus leading to a negative

Received March 1978; revised February 1980.

[1] Supported in part by the National Research Council of Canada Grant No. A4631.

AMS 1970 *subject classification.* Primary. 62G05.

Key words and phrases. Exact asymptotic behavior, density function, derivatives of a density, exact rate of convergence.

estimate for the density. But this is the price one pays to reduce the bias of the estimate. Kernels (1.1) for $p = 0$ and 1 have initially been used in estimation of a density and its first derivative by Johns and Van Ryzin (1972) for improving convergence rates for empirical Bayes two-action problems. A version of (1.3) for $p = 0$ and 1 has recently been used in Singh (1979b) to improve convergence rates for empirical Bayes squared error loss estimation problems.

2. The results. To describe our results here, we will make use of the following definition. We will say two sequences $\{a_n\}$ and $\{b_n\}$ are asymptotically equivalent, and write $a_n \sim b_n$, if $\lim_{n\to\infty}(a_n/b_n) = 1$. We recall the following definition.

DEFINITION 2.1. A point y on the real line R is a right Lebesgue-point of a real valued function g on R if $\epsilon^{-1}\int_y^{y+\epsilon} |g(t) - g(y)|\,dt \to 0$ as $\epsilon \downarrow 0$.

Notice that every right continuity point of a function is also a right Lebesgue-point of the function. The converse, however, is not true, e.g., take a function g which is 1 at every rational point and -1 elsewhere.

THEOREM 2.1. (Exact asymptotic expression of the bias). *If $f^{(r)}$ exists* a.e. *on* $[x, x + h)$ *and x is a right Lebesgue-point of $f^{(r)}$, then*

$$h^{-(r-p)}(E\hat{f}^{(p)}(x) - f^{(p)}(x)) \sim k_r f^{(r)}(x) \tag{2.1}$$

where

$$k_r = \frac{1}{r!}\int y^r K(y)\,dy.$$

PROOF. The proof follows from (3.5) of Singh (1977) since the second term on the right-hand side there is asymptotically equivalent to $f^{(r)}(x)$ times

$$k_r = \frac{h^{-r}}{(r-1)!}\int K(y)\int_x^{x+hy}(x + hy - t)^{r-1}\,dt\,dy. \qquad \square$$

THEOREM 2.2. (Exact asymptotic expression for MSE.) *Under the hypothesis of Theorem* 2.1,

$$\begin{aligned} &E(\hat{f}^{(p)}(x) - f^{(p)}(x))^2 \\ &\quad\sim n^{-1}\left[(n-1)(k_r h^{r-p} f^{(r)}(x))^2 + h^{-2p-1} f(x)\int K^2 - (f^{(p)}(x))^2 - 2h^{r-p} k_r f^{(p)}(x) f^{(r)}(x)\right]. \end{aligned} \tag{2.2}$$

PROOF. By (1.2) $EY_1^2(x)$ is h^{-2p-1} times

$$h^{-1}\int K^2\left(\frac{y-x}{h}\right)f(y)\,dy = h^{-1}\int_x^{x+h} K^2\left(\frac{y-x}{h}\right)f(y)\,dy$$

which is asymptotically equivalent to $f(x)\int K^2$. Thus, since by (1.3) $E\hat{f}^{(p)} = E(Y_1)$ and $\operatorname{Var}(\hat{f}^{(p)}) = n^{-1}\operatorname{Var}(Y_1)$, by Theorem 2.1 we get

$$\operatorname{Var}(\hat{f}^{(p)}(x)) \sim n^{-1}\left[\left\{h^{-2p-1} f(x)\int K^2\right\} - \{f^{(p)}(x) + h^{r-p} k_r f^{(r)}(x)\}^2\right]. \tag{2.3}$$

In view of Theorem 2.1, the proof of the theorem is now complete since MSE = $(\text{bias})^2$ + Var. $\square$

It is of interest to study the asymptotic behavior of $\operatorname{Cov}(\hat{f}^{(p)}(x_1), \hat{f}^{(p)}(x_2))$. Notice that for any two points $x_1 \neq x_2$ such that $h \leq |x_1 - x_2|$,

$$E(Y_1(x_1) \cdot Y_1(x_2)) = h^{-2p-1} \int K(t)K\left(t + \frac{x_1 - x_2}{h}\right) f(x_1 + ht)\, dt \equiv 0 \tag{2.4}$$

since K vanishes off (0, 1). Thus as an immediate consequence of Theorem 2.1 we have the following corollary.

Corollary 2.1. (Exact asymptotic expression for Cov.) *For any two points $x_1 \neq x_2$, let $f^{(r)}$ exist* a.e. *on $(x_1, x_1 + h) \cup [x_2, x_2 + h)$. If both points are right Lebesgue-points of $f^{(r)}$, then taking $h \leq |x_1 - x_2|$, we have*

$$\operatorname{Cov}(\hat{f}^{(p)}(x_1), \hat{f}^{(p)}(x_2)) \sim -n^{-1} \textstyle\prod_{i=1}^{2} \{f^{(p)}(x_i) + h^{r-p}k_r f^{(r)}(x_i)\}. \tag{2.5}$$

If x_1 ($\neq x_2$) is a right Lebesgue-point of f, then

$$h^{2p+1}E(Y_1(x_1) \cdot Y_1(x_2)) = \left\{h^{-1} \int_{x_1}^{x_1+h} K\left(\frac{y - x_1}{h}\right) K\left(\frac{y - x_2}{h}\right) (f(y) - f(x_1))\, dy\right\}$$
$$+ \left\{f(x_1) \int K(t)K\left(t + \frac{x_1 - x_2}{h}\right) dt\right\} = o(1)$$

since K is bounded and vanishes off (0, 1). Similarly $h^{2p+1}E(Y_1(x_1)) \cdot E(Y_1(x_2)) = o(1)$. Thus for every pair of points $x_1 \neq x_2$, one of which is right Lebesgue-point of f and for every $p = 0, 1, 2, \cdots$,

$$nh^{2p+1} \operatorname{Cov}(\hat{f}^{(p)}(x_1), \quad \hat{f}^{(p)}(x_2)) = o(1). \tag{2.6}$$

The relations (2.6) are useful in the study of asymptotic normality of the vector $(\hat{f}^{(p)}(x_1), \cdots, \hat{f}^{(p)}(x_m))$.

We will now investigate the exact asymptotic behavior of the error term $(\hat{f}^{(p)} - f^{(p)})$.

Theorem 2.3. *If $nh^{3r+1} \to 0$, and for every $t > 0$, $\sum_1^\infty \exp(-tnh^{2r+1}) < \infty$ with probability one, then $h^{-(r-p)}(\hat{f}^{(p)} - E\hat{f}^{(p)})$ is $o(1)$ as $n \to \infty$* w.p.1.

Proof. Let $T_j = K((X_j - x)/h) - EK((X_j - x)/h)$, $\bar{T} = n^{-1}\sum_1^n T_j$ and let M be the bound of K. In view of (1.2), (1.3) and the Borel-Cantelli lemma we only need to show that for every $\epsilon > 0$

$$\textstyle\sum_{n=1}^{\infty} P[h^{-(r+1)} |\bar{T}| > \epsilon] < \infty. \tag{2.7}$$

Notice that $T_1, \cdots T_n$ are i.i.d. centered random variables, each bounded by $2M$, and by (2.3), $\sigma^2 = n \operatorname{Var}(\bar{T}) \sim (hf(x) \int K^2) = ha$ (say). Consequently by Bernstein's inequality (see, for example, Bennett (1962)),

$$P[h^{-(r+1)} |\bar{T}| > \epsilon] \leq 2 \exp\left\{-\frac{nh^{2r+2}\epsilon^2}{2\sigma^2}\left(1 + \frac{2\epsilon Mh^{r+1}}{3\sigma^2}\right)^{-1}\right\}. \tag{2.8}$$

The right-hand side of (2.8) is $\sim \exp\{-nh^{2r+1}\epsilon^2/a\}$ since $\sigma^2 \sim ha$ and $nh^{3r+1} = o(1)$. Thus (2.7) follows by our second hypothesis on h. □

Theorem 2.4 shows that under its assumptions both the bias and the error of the estimators $\hat{f}^{(p)}$ are asymptotically equivalent. As a result, the following corollary is an immediate consequence of Theorems 2.1 and 2.4.

Corollary 2.2. (Exact asymptotic behavior of the error.) *Under the assumptions of Theorems 2.1 and 2.4,*

$$(\hat{f}^{(p)}(x) - (f^{(p)}(x)) \sim h^{(r-p)} f^{(r)}(x) k_r \quad \text{w.p.1.}$$

3. Concluding remarks. An optimal choice of h could be the one which minimizes the positive component of the exact asymptotic expression for the MSE given below. By Theorem 2.2,

$$\mathrm{MSE}(\hat{f}^{(p)}(x)) \sim \left[\left\{h^{r-p}k_r f^{(r)}(x)^2 + (nh^{2p+1})^{-1}f(x)\int K^2\right\} - n^{-1}\{f^{(p)}(x) + h^{(r-p)}k_r f^{(r)}(x)\}^2\right].$$

Thus an optimal choice of h could be

$$h = n^{-1/(1+2r)}\left\{\frac{(2p+1)f(x)\int K^2}{2(r-p)(k_r f^{(r)}(x))^2}\right\}^{1/(1+2r)}.$$

However, to make use of this value of h in the construction of the estimators $\hat{f}^{(p)}$, a good guess of the value of f and $f^{(r)}$ or of the magnitude of the ratio $f/(f^{(r)})^2$ is required beforehand. Notice that this optimal choice of h specialized to the case $p = 0$ coincides (up to the factors k_r and $\int K^2$) with that suggested by Parzen (1962) in the construction of estimators of a density.

REFERENCES

BENNETT, GEORGE (1962). Probability inequalities of sums of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

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

JOHNS, M. V., JR. AND VAN RYZIN, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. *Ann. Math. Statist.* **43** 934–943.

PARZEN, EMANUEL (1962). On the estimation of a probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

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

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

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

SINGH, R. S. (1977). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.* **5** 394–399.

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

SINGH, R. S. (1979b). Empirical Bayes estimation in Lebesgue exponential families with rates near the best possible rate. *Ann. Statist.* **7** 890–902.

DEPARTMENT OF MATHEMATICS & STATISTICS
UNIVERSITY OF GUELPH
ONTARIO N1G 2W1
CANADA