JOURNAL OF MULTIVARIATE ANALYSIS 9, 1–15 (1979)

Multivariate k-Nearest Neighbor Density Estimates

Y. P. MACK AND M. ROSENBLATT*

University of California, San Diego, La Jolla, California 92093

Communicated by P. R. Krishnaiah

Under appropriate assumptions, expressions describing the asymptotic behavior of the bias and variance of k-nearest neighbor density estimates with weight function w are obtained. The behavior of these estimates is compared with that of kernel estimates. Particular attention is paid to the properties of the estimates in the tail.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random p-vectors with *bounded continuous density* $f(x)$. Consider an estimate of $f(x)$ given by

$$f_n(x) = \frac{1}{nR_n{}^p} \sum_{j=1}^{n} w\left(\frac{x - X_j}{R_n}\right), \tag{1}$$

where

$R_n = R_n(x) =$ the Euclidean distance between x and the kth nearest neighbor of x among the X_j's,

w is a *bounded integrable* weight function with

$$\int w(u)\, du = 1,$$

and $k = k(n)$ is a sequence of positive integers such that $k \to \infty$, $k/n \to 0$ as $n \to \infty$.

In the past there has been extensive research on the properties of kernel estimates (See for example [2, 10, 11]) where the bandwidth (analogous to R_n) is deterministically specified. Recently there has been a good deal of interest in

Received October 25, 1978.

AMS 1970 subject classifications: Primary 62G05; Secondary 62G20.

Key words and phrases: Density estimates, k-nearest neighbor, kernel estimate, order statistics, variance, bias, tail behavior.

* Research supported in part by the Office of Naval Research Contract N00014-75-C-0428.

0047-259X/79/010001-15$02.00/0

nearest neighbor estimates. The basic idea of such an estimate is suggested in [5] and formalized in [7]. Results on consistency are found in [4] and [8] and asymptotic normality is discussed in [9]. Comments on the use of such estimates in classification are to be found in [3]. Some researchers are attracted by what they regard as the self-adjusting character of nearest neighbor estimates as contrasted with kernel estimates [1]. Most of the earlier papers deal only with a uniform weight function. One of the objects of this paper is to derive results on the asymptotic bias and variance of nearest neighbor estimates for a reasonably large class of weight functions. These results generalize those of Fukunaga and Hostetler [6] who obtained comparable results on heuristic grounds for a uniform weight function. A comparison is also made between nearest neighbor and kernel estimates showing that the bias of nearest neighbor estimates can be very large in the tail of the distribution in spite of their self-adjusting character.

We shall give a statement of the two basic theorems on variance and bias after introducing a little notation. Let $\| u \|$ denote the Euclidean norm of the p-vector u. Set $S_r = \{z : \| z - x \| < r\}$. Given a differentiable function $g(x)$, let $D_\alpha g$ denote the partial derivative of g with respect to x_α. If g is twice continuously differentiable and w has finite second order moments let

$$Q(g)(x) = \sum_{\alpha,\beta} \int u_\alpha u_\beta w(u)\, du\, D_\alpha D_\beta g(x). \tag{2}$$

The first theorem describes the asymptotic behavior of the variance.

THEOREM 1. *Let the density f be bounded. Assume that the weight function w is bounded with*

$$\int | u_\alpha | \, | w(u) | \, du < \infty, \qquad \alpha = 1,\ldots, p. \tag{3}$$

Further, let $1 - P(S_r) = O(r^{-\alpha})$ for some $\alpha > 0$ as $r \to \infty$. Consider a point x with $f(x) > 0$ and f continuously differentiable in a neighborhood of x. Then, if $k = k(n) \to \infty$, $k(n)/n \to 0$ as $n \to \infty$

$$\operatorname{Var}(f_n(x)) = \frac{f^2(x)}{k} \frac{\pi^{p/2}}{\Gamma\left(\frac{p+2}{2}\right)} \int w^2(u)\, du + o\left(\frac{1}{k}\right). \tag{4}$$

The second theorem is concerned with the asymptotic bias of nearest neighbor estimates.

THEOREM 2. *Let the density f be bounded. The weight function w is assumed to be bounded with*

$$\int \| u \|^2 \, | w(u) | \, du < \infty, \qquad \int u_\alpha w(u)\, du = 0, \quad \alpha = 1,\ldots, p. \tag{5}$$

Further, let x be a point with $f(x) > 0$ and f continuously differentiable up to second order in a neighborhood of x. Then

$$Ef_n(x) = f(x) + \frac{\left(\Gamma\left(\frac{p+2}{2}\right)\right)^{2/p}}{2\pi f(x)^{2/p}} Q(f)(x) \left(\frac{k}{n}\right)^{2/p}$$

$$+ \frac{\pi^{p/2}}{\Gamma\left(\frac{p+2}{2}\right)} \frac{f(x)}{k} \int_{\|u\|=1} w(u)\, d\Sigma(u)$$

$$+ o\left(\left(\frac{k}{n}\right)^{2/p} + \frac{1}{k}\right). \tag{6}$$

Σ is the uniform distribution on the surface of the p-sphere of unit radius.

Notice that Theorems 1 and 2 hold for a reasonably large class of density functions f with finite or infinite support. The results of Fukunaga and Hostetler are obtained from (4) and (6) by taking a uniform weight function w.

One can already see that the contribution of the (bias)^2 to the mean square error can overwhelm that of the variance in the tail of the distribution because of the factor $f(x)^{2/p}$ in the denominator of the second term on the right of (6). This does not happen in the case of a kernel estimate. Related questions will be discussed in greater detail in §4.

2. Preliminary Comments

Let us first consider the probability density $h(r)$ of the distance R_n between x and the kth nearest neighbor of x. Let $S_r = \{z: \| z - x \| < r\}$, $G(r) = P(S_r)$, and

$$G'(r) = \lim_{\delta \to 0} \frac{1}{\delta} \left[\int_{S_{r+\delta}} f(t)\, dt - \int_{S_r} f(t)\, dt\right] = \int_{\|x-t\|=r} f(t)\, d\sigma(t),$$

where P is the probability measure with density f, and σ is the surface area of the sphere $\| x - t \| = r$. Thus the density of R_n is

$$h(r) = n \binom{n-1}{k-1} G(r)^{k-1}(1 - G(r))^{n-k}\, G'(r). \tag{7}$$

The joint density of the kth nearest neighbor Q, the $k - 1$ observations $Y_1, \ldots, Y_{k-1}$ falling within the sphere about x whose radius is determined by x and Q, and the remaining $n - k$ observations $V_1, \ldots, V_{n-k}$ falling outside this sphere, is given as follows. Consider the joint density of $X_1, \ldots, X_n$. There are n choices possible for Q. Given that Q is chosen, there are $\binom{n-1}{k-1}$ possible choices for the $k-1$ observations falling within the sphere (and this determines the $n-k$ falling outside the sphere). The joint density of $Y_1, \ldots, Y_{k-1}$, $V_1, \ldots, V_{n-k}$ and Q is then

$$f(y_1, \ldots, y_{k-1}; v_1, \ldots, v_{n-k}; q) = n \binom{n-1}{k-1} \left\{ \prod_{j=1}^{k-1} f(y_j)\, \delta(y_j; S_r) \right\} \times \left\{ \prod_{l=1}^{n-k} f(v_l)\, \delta(v_l; (\bar{S}_r)^c) \right\} \cdot f(q) \tag{8}$$

where

$$\delta(z, A) = \begin{cases} 1 & \text{if } z \in A \\ 0 & \text{otherwise.} \end{cases}$$

$(\bar{S}_r)^c$ is the complement of $\bar{S}_r$, and $r = \| x - q \|$. Notice that if we first integrate the y_j's over S_r, the V_l's over $(\bar{S}_r)^c$, and then finally $f(q)$ over all of space we get one. If $f(q)$ is integrated over only $r = \| x - q \|$ we get $h(r)$. This implies that the conditional distribution of the Y_j's, the V_l's and Q given $R_n = r$ is

$$f(y_1, \ldots, y_{k-1}; v_1, \ldots, v_{n-k}; q \mid r) = \prod_{j=1}^{n-1} \left(\frac{f(y_j)}{G(r)} \right) \prod_{l=1}^{n-k} \left(\frac{f(v_l)}{1 - G(r)} \right) \cdot \left(\frac{f(q)}{G'(r)} \right),$$

so that the Y_j's, the V_l's and Q are conditionally independent given $R_n = r$ with respective marginal densities

$$\frac{f(y)}{G(r)}, \quad \frac{f(v)}{1 - G(r)}, \quad \text{and} \quad \frac{f(q)}{G'(r)},$$

$y \in S_r$, $v \in (\bar{S}_r)^c$, $q \in \{t : \| x - t \| = r\}$, where the conditional density of Q given R_n is to be integrated with respect to the surface measure on the sphere of radius r about x.

We are interested in computing moments of various functions of R_n. It is clear from what has been stated above that R_n has the same distribution as $G^{-1}(T)$, where T is the kth order statistic from an i.i.d. uniform (0, 1) sample of size n. If we just assume f is bounded and continuous we have

$$\begin{aligned} G(r) &= \int_{S_r} f(u)\, du = cf(x) r^p + \int_{S_r} [f(u) - f(x)]\, du \\ &= cf(w) r^p + o(r^p) \quad \text{as } r \downarrow 0, \quad \text{where } c = \frac{\pi^{p/2}}{\Gamma\left(\frac{p+2}{2}\right)}. \end{aligned}$$

Then if $t = G(r)$ it follows that when $f(x) > 0$

$$[G^{-1}(t)]^{\lambda} = r^{\lambda} = \left(\frac{t}{cf(x)}\right)^{\lambda/p} + o(t^{\lambda/p}). \tag{9}$$

In general, for ϕ a measurable function, $E(\phi(R_n))$ may not exist. However, since $o \leqslant G(r) \leqslant 1$, a sufficient condition for its existence is the requirement that

$$\int_0^{\infty} |\phi(r)|\, dG(r) < \infty. \tag{10}$$

Of particular interest to us in this discussion is the function

$$\phi(r) = \frac{r^{\lambda}}{G(r)^{\gamma}(1 - G(r))^{\beta}}, \tag{11}$$

where λ, γ, β are nonnegative integers. $E(\phi(R_n))$ exists for n sufficiently large under the assumption $(1 - G(r)) = 0(r^{-\zeta})$ as $r \to \infty$, $\zeta > o$, which implies (10), with ϕ given by (11).

With the change of variable $t = G(r)$, then

$$\begin{aligned} E(\phi(R_n)) &= n \binom{n-1}{k-1} \int_0^1 G^{-1}(t)^{\lambda} t^{k-1-\gamma} (1-t)^{n-k-\beta}\, dt \\ &= n \binom{n-1}{k-1} \int_0^1 \left(\left[\frac{t}{cf(x)}\right]^{\lambda/p} + o(t^{\lambda/p})\right) t^{k-1-\gamma} (1-t)^{n-k-\beta}\, dt. \end{aligned} \tag{12}$$

3. Proofs of the Theorems

We first derive the result on the asymptotic behavior of the variance of $f_n(x)$. Now

$$\mathrm{Var}(f_n(x)) = E[\mathrm{Var}(f_n(x) \mid R_n)] + \mathrm{Var}[E(f_n(x) \mid R_n)].$$

The following two propositions give us estimates of each of the terms on the right of the formula just written above.

Proposition 1. *Let the density f be bounded. Assume the weight function w is bounded and integrable. If f is continuous in a neighborhood of x and $k = k(n) \to \infty$, $k(n)/n \to 0$ as $n \to \infty$, then*

$$\begin{aligned} E[\mathrm{Var}(f_n(x) \mid R_n)] = {} & \frac{f^2(x)}{k} \frac{\pi^{p/2}}{\Gamma\left(\frac{p+2}{2}\right)} \int w^2(u)\, du \\ & - \frac{f^2(x)}{k} \left(\int_{\|u\| \leqslant 1} w(u)\, du\right)^2 + o\left(\frac{1}{k}\right). \end{aligned} \tag{13}$$

PROPOSITION 2. *Let the assumptions of Proposition* 1 *be satisfied. Let*

$$\int |u_\alpha| \, |w(u)| \, du < \infty, \qquad \alpha = 1, \ldots, p,$$

and $1 - P(S_r) = O(r^{-\alpha})$ *for some* $\alpha > 0$ *as* $r \to \infty$. *Then for* $f(x) > 0$ *and* f *continuously differentiable in a neighborhood of* x,

$$\operatorname{Var}[E(f_n(x) \mid R_n)] = \frac{f^2(x)}{k} \left(\int_{\|u\| \leqslant 1} w(u) \, du \right)^2 + o\left(\frac{1}{k}\right). \tag{13'}$$

Clearly Theorem 1 follows from Propositions 1 and 2. We first give the

Proof of Proposition 1

$$\begin{aligned} \operatorname{Var}(f_n(x) \mid R_n) = {} & \frac{k-1}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - Y}{R_n}\right) \Bigm| R_n \right) \\ & + \frac{1}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - Q}{R_n}\right) \Bigm| R_n \right) \\ & + \frac{n-k}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - V}{R_n}\right) \Bigm| R_n \right), \end{aligned} \tag{14}$$

where Y, Q, V have the distribution given in (8). Let $S(n) = \{y : \| x - y \| < R_n\}$. The first term on the right of (14) is

$$\begin{aligned} & \frac{k-1}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - Y}{R_n}\right) \Bigm| R_n \right) \\ & \quad = \frac{k-1}{n^2} \left[\frac{1}{R_n{}^p P(S(n))} \int_{\|u\| \leqslant 1} w^2(u) f(x - uR_n) \, du \right. \\ & \qquad \left. - \frac{1}{P(S(n))^2} \left(\int_{\|u\| \leqslant 1} w(u) f(x - uR_n) \, du \right)^2 \right]. \end{aligned}$$

Consider the first term on the right of this last equality. Its expectation

$$\begin{aligned} & E\left(\frac{k-1}{n^2} \frac{1}{R_n{}^p P(S(n))} \int_{\|u\| \leqslant 1} w^2(u) f(x - uR_n) \, du \right) \\ & \quad = \frac{k-1}{n^2} E\left(\frac{1}{R_n{}^p P(S(n))} \right) f(x) \int_{\|u\| \leqslant 1} w^2(u) \, du + E(A). \end{aligned} \tag{15}$$

By (12), the first term on the right of (15) equals

$$k^{-1} c f^2(x) \int_{\|u\| \leqslant 1} w^2(u) \, du + o(k^{-1}),$$

as $k \to \infty$ and we shall show that $E(A) = o(k^{-1})$. Notice that

$$| E(A)| \leqslant \frac{k}{n^2} \left\{ E\left(\frac{1}{R_n^{2p} P(S(n))^2}\right) \right.$$
$$\left. \times E\left(\int_{\|u\| \leqslant 1} w^2(u) \, |f(x - uR_n) - f(x)| \, du\right)^2 \right\}^{1/2},$$

by the Schwarz inequality. However

$$E(R_n^{-2p} P(S(n))^{-2}) = O\left(\frac{n}{k}\right)^{-4}, \tag{16}$$

and

$$\begin{aligned} E\left(\int_{\|u\| \leqslant 1} w^2(u) \, |f(x - uR_n) - f(x)| \, du\right)^2 \\ = \int_{\|u\|, \|v\| \leqslant 1} w^2(u) \, w^2(v) \, E\{|f(x - uR_n) - f(x)| \, |f(x - vR_n) - f(x)|\} \, du \, dv \end{aligned} \tag{17}$$

The expectation in the integrand on the right of (17) is bounded and converges to zero for all u, v since f is bounded and continuous. This remark implies that (17) converges to zero. Using (16), one can then see that $E(A) = o(1/k)$. A similar argument then shows that

$$E\left[\frac{k-1}{n^2} \frac{1}{P(S(n))^2} \left(\int_{\|u\| \leqslant 1} w(u) f(x - uR_n) \, du\right)^2\right]$$
$$= \frac{f(x)^2}{k} \left(\int_{\|u\| \leqslant 1} w(u) \, du\right) + o\left(\frac{1}{k}\right),$$

(using (12)). Another argument of the same type using (12) shows that

$$E\left[\frac{1}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - Q}{R_n}\right) \,\Big|\, R_n\right)\right] = o\left(\frac{1}{k}\right).$$

Again, using (12) and a similar argument one finds that

$$E\left[\frac{n-k}{n^2 R_n^{2p}} \operatorname{Var}\left(w\left(\frac{x - V}{R_n}\right) \,\Big|\, R_n\right)\right] = \frac{cf(x)^2}{k} \int_{\|u\| > 1} w^2(u) \, du + o\left(\frac{1}{k}\right).$$

Thus

$$E[\operatorname{Var}(f_n(x) \mid R_n)] = \frac{cf(x)^2}{k} \int w^2(u) \, du - \frac{f(x)^2}{k} \left(\int_{\|u\| \leqslant 1} w(u) \, du\right)^2 + o\left(\frac{1}{k}\right).$$

The following lemma gives useful estimates relating to the probability density of the kth order statistic and certain truncated moments. $1(A)$ denotes the indicator function of the set A.

LEMMA. *Let $T = T_{k,n}$ be the kth order statistic in a sample of size n from a $U[0, 1]$ distribution. If $k = k(n) \to \infty$, $k(n)/n \to 0$ as $n \to \infty$, then upper and lower bounds for the density function of T in the range $| T - (k-1)/(n-1)| < \alpha(k/n)$ with α small are given by*

$$\frac{1}{(2\pi)^{1/2}} \frac{n}{k^{1/2}} g_n(x) \exp\left\{\pm \frac{n^2}{k}\left(x - \frac{k-1}{n-1}\right)^2 \right. \\ \left. \times \left[\left|x - \frac{k-1}{n-1}\right| \frac{n}{k}\Big/\left(1 - \left|x - \frac{k-1}{n-1}\right| \frac{n}{k}\right)\right]\right\}, \tag{18}$$

where

$$g_n(x) = \exp\left\{-\frac{1}{2} \frac{(n-1)^3}{(k-1)(n-k)}\left(x - \frac{k-1}{n-1}\right)^2\right\}.$$

Further

$$E\{T^{-s}1(T < y)\} \leqslant \frac{k}{n} \frac{n}{k^{1/2}} \left(\frac{n}{k}\right)^s \{g_n(y)\}^{1/2}, \tag{19}$$

if $y = (1 - \alpha)\, k/n$ while

$$E\{T^{-s}1(T > y')\} \leqslant c\, \frac{k}{n} \frac{n}{k^{1/2}} \left(\frac{n}{k}\right)^s \{g_n(y')\}^{1/2} + \left(\frac{n}{k}\right)^s \exp(-\tfrac{1}{2}ck), \tag{20}$$

if $y' = (1 + \alpha)\, k/n$ and $c > 1$ is fixed but large.

The probability density of the kth order statistic in a sample of size n is

$$n \binom{n-1}{k-1} x^{k-1}(1 - x)^{n-k}, \qquad 0 \leqslant x \leqslant 1. \tag{21}$$

The maximum of the logarithm of the function $x^{k-1}(1 - x)^{n-k}$ is at

$$x_0 = \frac{k-1}{n-1}.$$

The second derivative of the logarithm at this point is

$$-\frac{(n-1)^3}{(k-1)(n-k)} \cong -\frac{n^2}{k},$$

and the sth derivative of the logarithm at x is

$$(s-1)!\left\{(-1)^{s+1}\frac{k-1}{x^s}-\frac{n-k}{(1-x)^s}\right\}.$$

The sth derivative at x_0 is bounded in absolute value by $(s-1)!\, n^s/k^{s-1}$ for large n. Using a Taylor expansion of the logarithm about x_0 and these bounds together with Stirling's approximation for $n\binom{n-1}{k-1}$ yields the upper and lower bounds of (18) in the range $|x-(k-1)/(n-1)| < \alpha\,(k/n)$ with α small. Because the density (21) is increasing for $x < (k-1)/(n-1)$

$$E\{T_{k,n}^{-s}1(T_{k,n} < y)\} \cong \left(\frac{n}{k}\right)^s E\{1(T_{k-s,n-s} < y)\}$$

$$\leqslant y\,\frac{n}{k^{1/2}}\left(\frac{n}{k}\right)^s \{g_n(y)\}^{1/2}, \qquad s = 0, 1, 2,$$

with $y = (1-\alpha)\,k/n$. The second tail moment inequality (20)

$$E\{T_{k,n}^{-s}1(T_{k,n} > y')\}$$

$$\cong \left(\frac{n}{k}\right)^s E\{1(T_{k-s,n-s} > y')\}$$

$$= \left(\frac{n}{k}\right)^s E\{1(y'' > T_{k-s,n-s} > y')\} + \left(\frac{n}{k}\right)^s E\{1(T_{k-s,n-s} \geqslant y'')\},$$

where $y' = (1+\alpha)\,k/n$ and $y'' = c(k/n)$ with c fixed but large. Since $(1-x)^{n-k} \leqslant \exp(-(n-k)\,x)$ for $0 \leqslant x \leqslant 1$

$$P(T_{n,k} \geqslant y'') \leqslant n\binom{n-1}{k-1}\int_{y''}^{\infty} x^{k-1}e^{-(n-k)x}\,dx$$

$$\cong \frac{1}{(k-1)!}\int_{(n-k)y''}^{\infty} u^{k-1}e^{-u}\,du$$

$$\cong \frac{1}{(k-1)!}\,[(n-k)y'']^{k-1}e^{-(n-k)y''}.$$

Thus, if c is fixed but much larger than one

$$E\{1(T_{k-s,n-s} > y'')\} \leqslant \exp(-\tfrac{1}{2}\,ck).$$

Also, since the density (21) is decreasing for $x > (k-1)/(n-1)$

$$E\{1(y'' > T_{k-s,n-s} > y')\} \leqslant c\,\frac{k}{n}\,\frac{n}{k^{1/2}}\left(\frac{n}{k}\right)^s \{g_n(y')\}^{1/2}.$$

We now give the

Proof of Proposition 2. The conditional expected value of $f_n(x)$ given R_n is

$$\begin{aligned} E(f_n(x) \mid R_n) &= \frac{k-1}{n} \frac{1}{R_n{}^p} \frac{1}{P(S(n))} \int_{S(n)} w\left(\frac{x-v}{R_n}\right) f(v)\, dv \\ &+ \frac{n-k}{n} \frac{1}{R_n{}^p} \frac{1}{1-P(S(n))} \int_{(S(n))^c} w\left(\frac{x-v}{R_n}\right) f(v)\, dv \\ &+ \frac{1}{n} \frac{1}{R_n{}^p} \int_{BS(n)} w\left(\frac{x-v}{R_n}\right) f(v)\, dv \Big/ \int_{BS(n)} f(v)\, dv \\ &= ① + ② + ③, \end{aligned} \tag{22}$$

where $(\bar{S}(n))^c$ is the complement of $\bar{S}(n)$ and $BS(n)$ is the boundary of $S(n)$. We shall make an estimate of Var(①). Similar estimates show that

$$\mathrm{Var}(②),\ \mathrm{Var}(③) = o\left(\frac{1}{k}\right).$$

All other contributions to Var(① + ② + ③), except for Var(①) are $o(1/k)$.

First note that

$$\begin{aligned} ① &= \frac{k-1}{n} \frac{1}{P(S(n))} \int_{\|u\| \leqslant 1} w(u) f(x - uR_n)\, du \\ &= \frac{k-1}{n} \frac{1}{T} \int_{\|u\| \leqslant 1} w(u) f(x - uG^{-1}(T))\, du, \end{aligned} \tag{23}$$

where T is the kth order statistic from a uniform distribution. In order to get sufficiently good estimates for various functions of T (such as, e.g., $1/T$, $T^{1/p}$, $G^{-1}(T)$) we require expansions in terms of $T - (k-1)/(n-1)$. Eventually certain moments are estimated in terms of contributions from the range $| T - (k-1)/(n-1)| < \alpha(k/n)$ and the complimentary range. From the assumptions made on w and f and the estimates (19) and (20), it follows that the contribution to $E(①)$ and $E(①^2)$ from the range $| T - (k-1)/(n-1)| \geqslant \alpha(k/n)$ is

$$O(e^{-(1/2)\alpha^2 k}).$$

We shall now consider the contribution of (23) to Var(①) from the range $| T - (k-1)/(n-1)| < \alpha\, k/n$. Now

$$\begin{aligned} \frac{k-1}{n}\frac{1}{T} &= \frac{k-1}{n}\frac{n-1}{k-1}\left(1 + \left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)\right)^{-1} \\ &= \frac{n-1}{n}\left\{1 - \left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)\right. \\ &\qquad \left. + O\left[\left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)\right]^2\right\}, \end{aligned} \tag{24}$$

while

$$\int_{\|u\|\leq 1} w(u)f\left(x - uG^{-1}\left(\frac{k-1}{n-1}\right) - u\left(G^{-1}(T) - G^{-1}\left(\frac{k-1}{n-1}\right)\right)\right) du$$

$$= \int_{\|u\|\leq 1} w(u)f\left(x - uG^{-1}\left(\frac{k-1}{n-1}\right)\right) du$$

$$- \int \sum_{\alpha} u_{\alpha}\, w(u) D_{\alpha} f\left(x - uG^{-1}\left(\frac{k-1}{n-1}\right)\right) du \left(G^{-1}(T) - G^{-1}\left(\frac{k-1}{n-1}\right)\right)$$

$$+ o\left(G^{-1}(T) - G^{-1}\left(\frac{k-1}{n-1}\right)\right).$$

It follows from (9) that

$$r = G^{-1}(t) = (cf(x))^{-1/p}\, t^{1/p} + o(t^{1/p}).$$

However,

$$t^{1/p} = \left(\frac{k-1}{n-1}\right)^{1/p}\left(1 + \left(\frac{k-1}{n-1}\right)^{-1}\left(t - \frac{k-1}{n-1}\right)\right)^{1/p}$$

$$= \left(\frac{k-1}{n-1}\right)^{1/p}$$

$$\times \left\{1 + \frac{1}{p}\left(\frac{k-1}{n-1}\right)^{-1}\left(t - \frac{k-1}{n-1}\right) + o\left(\left(\frac{k-1}{n-1}\right)^{-1}\left(t - \frac{k-1}{n-1}\right)\right)\right\}$$

so that

$$G^{-1}(T) - G^{-1}\left(\frac{k-1}{n-1}\right) = (cf(x))^{-1/p}\left(\frac{k-1}{n-1}\right)^{1/p}$$

$$\times \left\{\frac{1}{p}(1 + o(1))\left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)\right\}.$$

Neglecting the constant term we find that we have to estimate the second moment of

$$-\int_{\|u\|\leq 1} w(u)f\left(x - uG^{-1}\left(\frac{k-1}{n}\right)\right) du \left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)$$

$$- \int \sum u_{\alpha} w(u)\, D_{\alpha} f\left(x - uG^{-1}\left(\frac{k-1}{n-1}\right)\right) du\, \frac{n-1}{n}\,\frac{1}{p}\left(\frac{k-1}{n-1}\right)^{-1+1/p}$$

$$\cdot \left(T - \frac{k-1}{n-1}\right)(cf(x))^{-1/p}$$

$$+ o\left(\left(\frac{k-1}{n-1}\right)^{-1}\left(T - \frac{k-1}{n-1}\right)\right), \tag{25}$$

where it is understood that $|T-(k-1)/(n-1)| < \alpha(k/n)$. By using (24) we see that the second moment of the first term of (25) is to the first order

$$\left(\int_{\|u\|\leqslant 1} w(u)\,du\, f(x)\right)^2 \frac{1}{k},$$

while the second moments of the other terms are all

$$o\left(\frac{1}{k}\right).$$

We now give the

Proof of Theorem 2. From (22) it is clear that

$$Ef_n(x) = E(①) + E(②) + E(③).$$

Under the assumptions of the theorem

$$① = \frac{k-1}{nP(S(n))}\Big[f(x)\int_{\|u\|\leqslant 1} w(u)\,du$$

$$+ \frac{R_n^{\,2}}{2}\int_{\|u\|\leqslant 1}\sum_{\alpha,\beta} u_\alpha(D_\alpha D_\beta f)(x)\, u_\beta w(u)\,du + o(R_n^{\,2})\Big].$$

But then

$$E(①) = \frac{k-1}{n} f(x)\int_{\|u\|\leqslant 1} w(u)\,du\; E\left(\frac{1}{P(S(n))}\right)$$

$$+ \frac{k-1}{n}\,\frac{1}{2}\int_{\|u\|\leqslant 1}\sum_{\alpha,\beta} u_\alpha(D_\alpha D_\beta f)(x)\, u_\beta w(u)\,du\; E\left(\frac{R_n^{\,2}}{P(S(n))}\right)$$

$$+ o\left(E\left(\frac{k}{n}\,\frac{R_n^{\,2}}{P(S(n))}\right)\right)$$

$$= f(x)\int_{\|u\|\leqslant 1} w(u)\,du + \frac{1}{2(cf(x))^{2/p}}$$

$$\cdot \int_{\|u\|\leqslant 1}\sum_{\alpha,\beta} u_\alpha(D_\alpha D_\beta f)(x)\, u_\beta w(u)\,du \left(\frac{k}{n}\right)^{2/p}$$

$$+ o\left(\frac{k}{n}\right)^{2/p}.$$

Also

$$② = \frac{n-k}{n(1-P(S(n)))}\int_{\|u\|>1} w(u)\, f(x-uR_n)\,du,$$

and using an expansion similar to that used in (26) (as well as $\int_{\|u\|>s} |w(u)|\,du = o(s^{-2})$ as $s \to \infty$) and then taking expectations, one finds that

$$E(②) = f(x) \int_{\|u\|>1} w(u)\,du + \frac{1}{2(cf(x))^{2/p}}$$
$$\cdot \int_{\|u\|>1} \sum_{\alpha,\beta} u_\alpha (D_\alpha D_\beta f)(x)\, u_\beta w(u)\,du \left(\frac{k}{n}\right)^{2/p}$$
$$+ o\left(\frac{k}{n}\right)^{2/p}.$$

Further

$$③ = \frac{1}{nR_n{}^p} \int_{\|u\|=1} w(u) f(x - uR_n)\,d\sigma(u) \Big/ \int_{\|u\|=1} f(x - uR_n)\,d\sigma(u),$$

where σ denotes the surface area on the sphere of radius one. This implies that

$$E(③) = \frac{cf(x)}{k} \int_{\|u\|=1} w(u)\,d\Sigma(u) + o\left(\frac{1}{k}\right).$$

4. Comparison of Nearest Neighbor and Kernel Estimates

Consider the kernel density function estimate $\tilde{f}_n(x)$

$$\tilde{f}_n(x) = \frac{1}{nb(n)^p} \sum_{j=1}^{n} w\left(\frac{x - X_j}{b(n)}\right),$$

with bandwith $b(n) \downarrow o$, $nb(n)^p \to \infty$. Under the assumptions of Theorem 2

$$E\tilde{f}_n(x) - f(x) = \tfrac{1}{2}\, b(n)^2 Q(f)(x) + o(b(n)^2),$$

and

$$\mathrm{Var}(\tilde{f}_n(x)) = \frac{f(x)}{nb(n)^p} \int w^2(u)\,du + o\left(\frac{1}{nb(n)^p}\right),$$

as $n \to \infty$ (see [2, 10, 11]). If we consider setting $b(n) = Cn^{-\alpha}$, $\alpha > 0$, for a kernel estimate, the optimal rate of decay for the mean square error of the estimate is obtained when $\alpha = (p + 4)^{-1}$ and then the rate of decay is $\cong n^{-4/(4+p)}$. In the case of a nearest neighbor estimate, if one sets $k = C'n^\beta$, $\beta > 0$, the optimal rate of decay of the mean square error of the estimate is obtained with $\beta = (4)/(p + 4)$ and the rate of decay is then again $\cong n^{-4/(4+p)}$. The constant multiplying the rate of decay of the mean square error in the case of a kernel estimate is minimized by taking

$$C = \left\{ pf(x) \int w^2(u)\,du[b(n)^4\, Q(f)(x)^2]^{-1} \right\}^{1/(p+4)},$$

and it is found to be

$$\left[f(x)\int w^2(u)\,du\right]^{4/(4+p)}\left[\frac{1}{2}\,Q(f)(x)\right]^{2p/(4+p)}\left\{\left(\frac{4}{p}\right)^{p/(4+p)}+\left(\frac{p}{4}\right)^{4/(p+4)}\right\}. \quad (27)$$

The constant multiplying the rate of decay of the mean square error of the nearest neighbor estimate is minimized by taking

$$C'=\left(\frac{\pi}{\Gamma\left(\frac{p+2}{2}\right)}\right)^{p/2}\left(f(x)^{2+4/p}\int w^2(u)\,du/(Q(f)(x))^2\right)^{p/(4+p)},$$

and is found to be (27) again. We then find that to the first order the mean square error is the same for both estimates. However, this is something of an illusion since we don't know $f(x)$ and $Q(f)(x)$ and if we did would not have recourse to density estimates.

Notice that when $f(x)$ is large the variance of the kernel estimate appears to be smaller than that of the nearest neighbor estimate while the bias of the kernel estimate appears to be larger than that of the nearest neighbor estimate. Exactly the opposite appears to happen when $f(x)$ is small. In this discussion it is assumed that $b(n)$ and $(k/n)^{1/p}$ are comparable. So even though the variance of the nearest neighbor estimate is appreciably smaller than the variance of the kernel estimate, the bias of the nearest neighbor estimate

$$\frac{\Gamma\left(\frac{p+2}{2}\right)^{2/p}}{2\pi f(x)^{2/p}}\,Q(f)(x)\left(\frac{k}{n}\right)^{2/p}$$

looks as if it can be much larger than that of the kernel estimate

$$\tfrac{1}{2}\,b(n)^2\,Q(f)(x),$$

because of the factor $f(x)^{2/p}$ in the denominator. We shall see that in the case of a large class of densities, the bias is the main term when $f(x)$ is small and that contrary to hopes expressed (see [1, 9, 12]) the nearest neighbor estimate may be much worse than the kernel. This is illustrated by looking at densities having a simple exponential form or inverse polynomial decay in the tail. Just the one-dimensional case is considered. We can assume that $b(n)$, $(k/n) \sim n^{-1/4}$. First consider $f(x)=ce^{-ax^b}$ with $c, a, b>0$. Then

$$\frac{f''(x)}{f^2(x)}=c^{-1}e^{ax^b}\{a^2b^2x^{2(b-1)}-ab(b-1)x^{b-2}\}.$$

Also if $f(x)=cx^{-b}$, $c, b>o$, for x large, then $f''(x)/f^2(x)=(b(b+1)/c)\,x^{b-2}$, and this expression becomes large as $x\to\infty$ if $b>2$.

References

[1] Breiman, L., Meisel, W., and Purcell, E. (1977). Variable kernel estimates of multivariate densities and their calibration. *Technometrics* **19** 135–144.

[2] Cacoulos, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 179–189.

[3] Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* **13** 21–29.

[4] Devroye, L. P., and Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5** 536–540.

[5] Fix, E., and Hodges, J. L. (1951). "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties." Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

[6] Fukunaga, K., and Hostetler, L. D. (1973). Optimization of k-nearest neighbor density estimates. *IEEE Trans. Inform. Theory* **19** 320–326.

[7] Loftsgaarden, D. O., and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36** 1049–1051.

[8] Moore, D. S., and Yackel, J. W. (1977). Consistency properties of nearest neighbor density estimates. *Ann. Statist.* **5** 143–154.

[9] Moore, D. S., and Yackel, J. W. (1977). Large sample properties of nearest neighbor density function estimators. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, Eds.). Academic Press, New York.

[10] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[11] Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

[12] Wagner, T. J. (1975). Nonparametric estimates of probability density. *IEEE Trans. Inform. Theory* **IT-21** 438–440.