On a New Class of Density Estimators

Carlos C. Rodríguez
Department of Mathematics and Statistics
State University of New York at Albany
E-Mail: carlos@omega.albany.edu

December 21, 1992

Contents

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random p-vectors with unknown bounded continuous density f with respect to the Lebesgue measure λ on $\mathbb{R}^p$. Consider the classical k-nearest neighbor density estimator of $f(z)$ (see Devroye's book [1] or the papers [3], [6]) given by

$$h_n(z) = \frac{k/n}{\lambda\left(S(R(k))\right)} \tag{1}$$

where

$$S(r) = \{x \in \mathbb{R}^p : \|x - z\| \leq r\}$$

is the p-sphere about z of radius r. We consider $\|\cdot\|$ to be an arbitrary norm on $\mathbb{R}^p$. Also,

$R(k) =$ the $\|\cdot\|$-distance from z to the kth nearest neighbor of z among the X_j's.

and $k = k(n)$ is a sequence of positive integers such that:

$$k \to \infty,\ k/n \to 0 \text{ as } n \to \infty. \tag{2}$$

To simplify the notation we shall denote by λ_k the denominator of equation (1).

Equation (1) is just the proportion of points inside $S(R(k))$ divided by its volume λ_k and therefore it is a natural approximation to the Radon-Nikodyn derivative that defines $f(z)$. On the other hand, if the $||\cdot||$ used is regular in the sense that

$$\lambda(S(r)) = \lambda(S(1))r^p \equiv \beta r^p \quad \text{for all } r > 0. \tag{3}$$

the classical kernel density estimator can be written as:

$$g_n(z) = \frac{\int \beta K\left(\frac{z-x}{\mu}\right) dF_n(x)}{\lambda(S(\mu))} \tag{4}$$

where F_n is the empirical distribution, the kernel K is a bounded density on $\mathbb{R}^p$ and $\mu = \mu(n)$ is such that $k = [n\mu^p]$ satisfies (2).

The k-nn (1) and the kernel (4) methods can be seen as two extremes of a continuum. Both, $h_n(z)$ and $g_n(z)$ estimate $f(z)$ as *probability-mass-per-unit-volume*. The k-nn fixes the mass to the deterministic value of k/n and lets the volume λ_k to be stochastic, while the kernel method fixes the volume $\lambda(S(\mu))$ and lets the mass to be random (this is most evident when K is the uniform density on the unit sphere centered at 0). The kernel K acts as a weight function allocating more mass to the points that are more securely inside the sphere $S(\mu)$ (i.e. those points that are closer to the center, z).

The density estimators introduced bellow (5) and (8) fill up the gap between (1) and (4) in two different ways. First, they estimate the density as *mass-per-unit-volume* where both the *mass* and the *volume* are stochastic (e.g. functions of the observations). Secondly, they estimate *mass* and *volume* by smoothing the contribution of each point in the sample with different smoothing functions for the numerator and denominator. The novelty of these classes of estimators resides precisely in this **double** smoothing of *mass* and *volume*. Thus, equation (5) can be read as: "empirical *mass* F_n smoothed with the function K over empirical *volume* h^p smoothed with the function ω". Equation (8) has a similar interpretation but with a higher degree of symmetry between numerator and denominator.

Several authors have tried to combine the k-nn and kernel methods by replacing μ in the kernel (4) with the stochastic radius $R(k)$ appearing in the k-nn (1) (e.g. see [8], [7]). However, the only attempt at smoothing empirical volumes appears to be the Maximum Entropy Histograms (see [9], [10]) and only for $p = 1$ and uniform smoothing functions K and ω. This simple uniform smoothing of the denominators of (1) reduces the asymptotic variance and motivated the definition of (5) and (8).

We consider the class of density estimators that can be written as:

$$f_n(z) = \frac{\int K\left(\frac{z-x}{\mu}\right) dF_n(x)}{c^{-1}\int_0^1 h^p(u)\omega(u)du} \tag{5}$$

where ω is a density on $[0,1]$ with mean c i.e. $\forall u \in [0,1]$

$$\omega(u) \geq 0, \quad \int_0^1 \omega(u)du = 1 \quad \text{and} \quad \int_0^1 u\omega(u)du = c. \tag{6}$$

In (5) μ and $h(u)$ may be stochastic but they are assumed to be functions of the sample size n, satisfying $\forall u \in (0,1)$ and as $n \to \infty$ the following four conditions:

i. $h^p(u) = \mu^p u + o(\mu^p)$

ii. $h(u) \to 0$

iii. $nh^p(u) \to \infty$

(7)

iv. The sequence $\{h^p(u)/\mu^p\}_n$ is either monotone **OR** bounded above by a function $s(u)$ with $s\omega$ integrable on $(0,1)$ **OR** the little o in (i) holds uniformly in $u \in (0,1)$.

The little o and the limits above should be understood as almost sure when necessary.

The following class of density estimators includes, as special cases, several well known estimates (see [2] pp. 192-194 and references therein) and it is similar to (5):

$$\hat{f}_n(z) = \frac{\int_0^1 \omega(u)K\left(\frac{z - X_{[nu]+1}}{h(u)}\right) du}{\int_0^1 \omega(u)h^p(u)du} \tag{8}$$

The class defined above shows a nicer symmetry between the numerator and the denominator than the class (5). Moreover, density estimators defined by (8) have the very desirable property of integrating to one for all n whenever the function $h(u)$ is taken independent of z.

Our main concern in this paper is the class (5). The following are some members of this family:

Case I: (Classical kernel)
take $h^p(u) = \mu^p u$ and μ deterministic.

Case II: (Classical k-nn)
$K(x) = \beta^{-1} I(\|x\| \leq 1)$ and $h^p(u) = R^p(k)u$ with $\mu = R(k)$.Where the indicator function $I(A)$ takes the value 1 when the proposition A is true and 0 otherwise.

Case III: (Deterministic wights)

For a given number $b \geq 1$ (we shall assume, without loss of generality that bk is an integer) define:

$$h^p(u) = b^{-1} \sum_{i=1}^{bk} R^p(i) I \left(\frac{i-1}{bk} < u \leq \frac{i}{bk} \right) \tag{9}$$

Substituting (9) back in (5) we obtain the estimator:

$$f_n(z) = \frac{cb \sum_{i=1}^{n} K \left(\frac{z - x_i}{R(k)} \right)}{n \sum_{i=1}^{bk} \omega_i R^p(i)} \tag{10}$$

where

$$\omega_i = \int_{(i-1)/bk}^{i/bk} \omega(u) du.$$

Case IV: (Random weights)

For $r > 0$ define $n(r) =$ "number of observations with $||z - x_j|| \leq r$". Let $H_n(r) = n(r)/n$ i.e. the empirical distribution of $||X - z||$ when the vector X has density f. Notice that $n(r)$ is binomial$(n, H(r))$ where H is the distribution function of $||X - z||$. Define for $u \in [0, 1]$ $h(u) = \inf\{r : H_n(r) > uH_n(\mu)\}$. Notice that $h(u) = H_n^{-1}(uH_n(\mu))$ has the same form of (9) but now instead of bk we use $n(\mu)$ obtaining:

$$f_n(z) = \frac{c \sum_{i=1}^{n} K \left(\frac{z - x_i}{\mu} \right)}{n \sum_{i=1}^{n(\mu)} \omega_i R^p(i)} \tag{11}$$

Guided by these examples we can say that (5) estimates the density of probability at z by estimating *mass* as a functional of the empirical distribution of X, F_n and estimating *volume* as a functional of the empirical quantile function of $||X - z||$, H_n^{-1}. Hence, it is expected that stochastic convergence properties for these estimators could be obtained from related properties for the empirical distribution and quantile processes.

In section 2 we show that estimators (5) inherit the strong consistency property of kernel estimators and that under some conditions they are also asymptotically gaussian at the usual optimal rates. We also verify that the estimators (10) and (11) satisfy (7) and thus, they are strongly consistent estimators of $f(z)$.

In section 3 we present the main result of this paper. We prove that, under second order conditions on f, estimators (10) are asymptotically gaussian at the usual optimal rates and we show explicit formulas for the asymptotic bias and variance. These formulas differ from the corresponding expressions for the kernel and k-nn in that they depend on the smoothing function for volumes ω through the integrals:

$$c^{-2}\int_0^1 \left\{\int_y^1 \omega(x)dx\right\}^2 dy \quad \text{and} \quad c^{-1}\int_0^1 u^{1+2/p}\omega(u)du. \tag{12}$$

Section 4 is devoted to the computation of the optimal weight function ω under first order conditions on f. It is also shown there that the estimators (10) strongly dominate the classical kernel and the generalized k-nn (see [7]). We find that no matter what smoothness parameter is used in the generalized k-nn (or classical kernel) estimator with kernel K, there is always an estimator from the class (10) with the same kernel with a strictly smaller *asymptotic mean squared error* (AMSE). This holds uniformly in $f \in \mathcal{C}^2, z \in \mathbb{R}^p, \beta, K, c, p, ||\cdot||$, and ω and by a correct choice of ω the ratio of the AMSEs can be made to increase without limit.

In section 5 we optimize the $||\cdot||$ for the sub class of (10) of k-nn type density estimators, i.e. obtained when K is the uniform density on the unit $||\cdot||$-sphere about 0. We find that associated to this k-nn class of estimators there is a Riemannian manifold with the property that if the estimators are computed based on its metric, the AMSE is minimized. This suggests changing the $||\cdot||$ adaptively at each point z.

2 Consistency and Asymptotic Normality Inherited from the Kernel

The following two propositions are about the general estimators f_n of the form (5) satisfying (6) and (7).

Proposition 2.1 (Strong consistency of f_n) *If $n\mu^p/\log\log\, n \to \infty$ a.s. as $n \to \infty$ then $\forall z \in \mathbb{R}^p$ we have $f_n(z) \to f(z)$ almost surely.*

PROOF. Let $g_n(z)$ be as in (4). Then, we can write

$$f_n(z) = g_n(z)/\left\{\int_0^1 \frac{h^p(u)}{c\mu^p}\omega(u)du\right\} \equiv g_n(z)/\tilde{h}_n(z) \tag{13}$$

The numerator $g_n(z)$ goes almost surely to $f(z)$ (see [2] p.149 theorem 2B). By (7) part (i) and the dominated or monotone convergence theorem (depending

on what holds in (7)(iv)) the denominator goes (a.s.) to $c^{-1}\int_0^1 u\omega(u)du = 1$.
□

Let us define:

$$
\begin{aligned}
k &= [n\mu^p] && (14)\\
Z_n^K &= \sqrt{k}\,(g_n(z) - f(z)) && (15)\\
Z_n^\omega &= \sqrt{k}\left(\tilde{h}_n(z) - 1\right). && (16)
\end{aligned}
$$

The following simple result shows that the estimators f_n inherit the asymptotic normality property from the corresponding property for kernel estimators g_n, if the sequence of stochastic processes $\left\{\sqrt{k}(\mu^{-p}h^p(u) - u); 0 \le u \le 1\right\}_n$ is asymptotically gaussian and uncorrelated with g_n.

Proposition 2.2 (Asymptotic normality of f_n) *If as $n \to \infty$ the random variables Z_n^K and Z_n^ω converge in law to gaussian r.v.s $N(B_K(z), V_K(z))$ and $N(B_\omega(z), V_\omega(z))$ resp. with $\lim_{n\to\infty} cov(Z_n^K, Z_n^\omega) = 0$ then:*

$$\sqrt{k}\,(f_n(z) - f(z)) \stackrel{\mathcal{L}aw}{\to} N\left(B(z), V(z)\right) \tag{17}$$

where

$$
\begin{aligned}
B(z) &= B_K(z) - B_\omega(z)f(z) && (18)\\
V(z) &= V_K(z) + V_\omega(z)f^2(z) && (19)
\end{aligned}
$$

Moreover, if

$$\left\{\sqrt{k}\,(h^p(u)/\mu^p - u)\,; 0 \le u \le 1\right\} \stackrel{\mathcal{D}}{\to} \{B_h(u) + V_h(u)\mathcal{B}(u); 0 \le u \le 1\} \tag{20}$$

where $\{\mathcal{B}(u); 0 \le u \le 1\}$ is a separable gaussian process with $E\mathcal{B}(u) = 0$ and covariance function $E\mathcal{B}(u_1)\mathcal{B}(u_2) = \mathcal{V}(u_1, u_2)$. Then Z_n^ω is asymptotically gaussian with

$$
\begin{aligned}
B_\omega(z) &= c^{-1}\int_0^1 B_h(u)du && (21)\\
V_\omega(z) &= c^{-2}\int_0^1\int_0^1 V_h(u)V_h(v)\mathcal{V}(u,v)\omega(u)\omega(v)dudv. && (22)
\end{aligned}
$$

In particular if $\mathcal{V}(u,v) = u \wedge v - uv$ (i.e. the standard Brownian Bridge) then

$$V_\omega(z) = \frac{c^{-2}}{2}\int_0^1 V_h(u)u(1-u)\omega(u)du. \tag{23}$$

PROOF. We have

$$\sqrt{k}\,(f_n(z) - f(z)) = \left\{Z_n^K - Z_n^\omega f(z)\right\}/\tilde{h}_n(z)$$

hence (17) follows from Slutzky's theorem applied to the previous equation since $\tilde{h}(z) \to 1$ (a.s.). □

The following two propositions are consequences of proposition 2.1.

Corollary 2.1 (Strong consistency of Case III) *Let $f_n(z)$ be given by (10). If $k/\log\log n \to \infty$ and $k/n \to 0$ as $n \to \infty$ then $f_n(z) \to f(z)$ a.s.*

Corollary 2.2 (Strong consistency of Case IV) *Let $f_n(z)$ be given by (11) and let $k = n\mu^p$. If $k/\log\log n \to \infty$ and $k/n \to 0$ as $n \to \infty$ then $f_n(z) \to f(z)$ a.s.*

For completeness and future reference we first prove the strong consistency of the classical k-nn estimator (1) under no assumptions on f.

Define,

$$H(r) = P\left[\|X - z\| \le r\right] = \int_{S(r)} f(x)\lambda(dx) \tag{24}$$

then $R(1) < R(2) < \ldots < R(n)$ are the order statistics of n iid random variables with parent distribution $H(r)$. We shall use Lemma 1 from [8] which is proved in [4].

Lemma 2.1 (Kiefer) *Let z_n be a sample α_n-tile from n iid random variables uniformly distributed on $(0,1)$. If $\alpha_n \to 0$ and $n\alpha_n/\log\log n \to \infty$ as $n \to \infty$, then $z_n/\alpha_n \to 1$ a.s.*

We have that $H(R(1)) < H(R(2)) < \ldots < H(R(n))$ are the order statistics of n iid random variables uniformly distributed on $(0,1)$ with k/n-tile given by $H(R(k)) \equiv U_{k:n}$.

Lemma 2.2 (Strong consistency of k-nn) *If $k/\log\log n \to \infty$ and $k/n \to 0$ as $n \to \infty$ then*

$$\frac{k/n}{\lambda_k} \to f(z) \quad a.s. \tag{25}$$

PROOF. Write

$$\frac{k/n}{\lambda_k} = \left\{\frac{k/n}{U_{k:n}}\right\} \cdot \left\{\frac{\int_{S(R(k))} f(x)\lambda(dx)}{\lambda_k}\right\}$$

and apply Lemma 2.1 to the first term and the Lebesgue density theorem to the second. □

PROOF. (Corollary 2.1) Notice that by proposition 2.1 we only need to verify (7)i) and (7)iv). From the definition of $h(u)$ we have:

$$\frac{h^p(u)}{\mu^p} = \frac{R^p(i_n(u))}{bR^p(k)}$$

where $i_n(u) = \min\{i : i \geq bku\}$. Hence, $|i_n(u)/bk - u| < (bk)^{-1}$, $i_n(u)/n \to 0$ and, $i_n(u)/\log\log n \to \infty$. Therefore, applying Lemma 2.2 twice

$$\frac{h^p(u)}{\mu^p} = \left\{\frac{k/n}{\beta R^p(k)}\right\}\left\{\frac{i_n(u)/n}{\beta R^p(i_n(u))}\right\}^{-1}\frac{i_n(u)}{bk} = u + o(n^0) \qquad (26)$$

holds a.s. as $n \to \infty$. (Notice that we only need $\lambda_k = \beta R^p(k) + o(n^0)$). This shows (7)i). To verify (7)iv) notice that from (26) we can always take n, large enough, so that the RHS is bounded by 2 and we obtain (7)iv) with $s(u) \equiv 2$. □

PROOF. (Corollary 2.2) We only need to verify (7)i) and (7)iv). We have $\forall u \in (0,1)$

$$\frac{h^p(u)}{\mu^p} = \frac{R^p(un(\mu))}{\mu^p} = \left\{\frac{H_n^{-1}(uH_n(\mu))}{\mu}\right\}^p \leq 1. \qquad (27)$$

which shows (7)iv) with $s(u) \equiv 1$. By the strong uniform consistency of the empirical distribution H_n to the true distribution H we can write

$$\frac{n(\mu)}{n\mu^p} = \frac{H(\mu)}{\mu^p} + o(n^0)$$

and by the Lebesgue density theorem

$$\frac{H(u)}{\mu^p} = \beta\int_{S(\mu)} f(x)\lambda(dx)/\lambda(S(\mu)) \to \beta f(z)$$

holds as $\mu \to 0$ for almost all $z \in \mathbb{R}^p$. Thus,

$$\frac{n(\mu)}{n\mu^p} = \beta f(z) + o(n^0)$$

and from (27) and Lemma 2.2 we have,

$$\text{as-}\lim_{n\to\infty}\frac{h^p(u)}{\mu^p} = \text{as-}\lim_{n\to\infty}\left\{\frac{R^p(u\beta f(z)n\mu^p)}{u\beta f(z)n\mu^p/n}\right\}\cdot u\beta f(z) = u$$

since $k = u\beta f(z)n\mu^p$ satisfies $k/\log\log n \to \infty$ and $k/n \to 0$ as $n \to \infty$. □

3 Asymptotic Normality of Estimator (10)

The main result of this paper is the asymptotic normality property of estimators defined by equation (10) above. Before enunciating the theorem we state the conditions on the $||\cdot||$ and the kernel.

Conditions on the Norm

Let $S_0(r) = \{x \in \mathbb{R}^p : ||x|| \leq r\}$. We assume that the $||\cdot||$ satisfies:

$$\begin{aligned} &\text{i. } \exists \beta > 0 \;\; \lambda(S(r)) = \beta r^p + o(r^{p+2}) \text{ as } r \to 0 \\ &\text{ii. } ||\textstyle\int_{S_0(r)} y\lambda(dy)|| = o(r^{p+2}) \text{ as } r \to 0. \end{aligned} \tag{N}$$

Notice that these are only mild conditions on the $||\cdot||$. All the Hölder norms (including the norm of the uniform convergence) satisfy (N) with the $o(r^{p+2}) \equiv 0$.

Conditions on the Kernel

Let $K : \mathbb{R}^p \to \mathbb{R}^+$ be a bounded function satisfying

$$\begin{aligned} &\text{i. } \textstyle\int_{S_0(1)} K(y)\lambda(dy) = 1 \quad \text{(K is a density on } S_0(1)) \\ &\text{ii. } ||\textstyle\int_{S_0(1)} yK(y)\lambda(dy)|| = 0 \quad \text{(K is symmetric about 0).} \end{aligned} \tag{K}$$

We state the theorem under the usual second-order conditions on the density f.

Theorem 3.1 (Asymptotic normality of (10)) *Let $f_n(z)$ be defined by equation (10) with the norm and the kernel satisfying (N) and (K) respectively. If*

$$\lim_{n\to\infty} n^{\frac{-4}{p+4}} k = a > 0 \tag{28}$$

and the second order partial derivatives of f exist and they are continuous at z with $f(z) > 0$ then as $n \to \infty$

$$\sqrt{k}\left(f_n(z) - f(z)\right) \xrightarrow{\mathcal{L}aw} N\left(B(z), V(z)\right),$$

where the asymptotic bias $B(z)$ is given by

$$B(z) = \frac{a^{\frac{p+4}{2p}}}{2\left[\beta f(z)\right]^{2/p}} \left\{\eta(z) + \frac{b^{2/p}}{\beta c}\gamma(z)\delta\right\} \tag{29}$$

with

$$\gamma(z) = \int_{S_0(1)} x' H_2(z) x \lambda(dx) \tag{30}$$

$$\eta(z) = \int_{S_0(1)} x' H_2(z) x \left[K(x) - \beta^{-1}\right] \lambda(dx) \tag{31}$$

and where $H_2(z)$ denotes the Hessian of f evaluated at z. The asymptotic variance $V(z)$ is given by

$$V(z) = \beta f^2(z) \int_{S_0(1)} K^2(x)\lambda(dx) - (1-\alpha) f^2(z) \tag{32}$$

with α and δ being functionals of the weight function ω, given by

$$\alpha = c^{-2} b^{-1} \int_0^1 \left\{ \int_y^1 \omega(x) dx \right\}^2 dy \tag{33}$$

$$\delta = \int_0^1 u^{1+\frac{2}{p}} \omega(u) du. \tag{34}$$

The rate of convergence to normality in theorem 3.1 can be improved by strengthening the smoothing of f without assuming differentiability of order higher than 2. Let us denote by (N') the norm conditions (N) with $o(r^{p+2}) \equiv 0$. We shall use the following standard notation: For $j = (j_1, \ldots, j_p) \in \mathbb{N}^p$ define $|j| = j_1 + j_2 + \cdots + j_p$ and $j! = j_1! \cdots j_p!$ and for $x \in \mathbb{R}^p$, $x^j = x_1^{j_1} x_2^{j_2} \cdots x_p^{j_p}$ and $f^{(j)}(x) = \partial^{|j|} f(x) / \partial x_1^{j_1} \cdots \partial x_p^{j_p}$. A small modification to the proof of theorem 3.1 will produce:

Theorem 3.2 *Let $f_n(z)$ be defined by equation (10) with the kernel satisfying (K) and the norm satisfying (N'). If $f(z) > 0$ and the second mixed partial derivatives of f at z have continuous non decreasing modulus of continuity ξ with $\xi(0) = 0$, i.e. there exists a neighborhood $\mathcal{V}$ of z with*

$$\left| f^{(j)}(x) - f^{(j)}(z) \right| \leq \xi(\|x - z\|) \quad \forall x \in \mathcal{V}, |j| = 2, \tag{35}$$

and

$$\left(\frac{k}{n}\right)^{\dfrac{p+4}{2p}} \xi\left(\left(\frac{k}{n}\right)^{1/p} \right) = o\left(n^{-1/2}\right). \tag{36}$$

Then as $n \to \infty$

$$\sqrt{k} \left(f_n(z) - f(z) - \left(\frac{k}{n}\right)^{2/p} B'(z) \right) \stackrel{\mathcal{L}aw}{\rightarrow} N(0, V(z)) \tag{37}$$

where the asymptotic variance $V(z)$ is given by (32) and

$$B'(z) = \frac{1}{2}\eta(z) + \frac{b^{2/p}}{2c\beta}\gamma(z)\delta \tag{38}$$

and $\eta(z), \gamma(z)$ and δ are defined as in theorem 3.1 (see equations (31), (30) and (34) above). In particular if $\xi(u) = Au^\tau$ with $0 < \tau < 1$ then (36) is equivalent to

$$\lim_{n \to \infty} k \exp\left\{ \frac{-2(\tau+2)}{p + 2(\tau+2)} \log n \right\} = 0.$$

Therefore, the rate of convergence to normality improves from $n^{\frac{-4}{p+4}}$ *of theorem 3.1 (obtained when* $\tau \to 0$*) to* $n^{\frac{-6}{p+6}}$ *approached when* $\tau \to 1$.

PROOF. (of theorem 3.1) The schema of the proof follows [8] theorem 3, i.e. we show that we can write

$$\sqrt{k}\,(f_n(z) - f(z)) = A_n + B_n \tag{39}$$

where

$$A_n \stackrel{\mathcal{L}aw}{\to} N\left(0, \beta f^2(z) \int K^2(y)\lambda(dy) - f^2(z)\right) \tag{40}$$

and

$$B_n \stackrel{\mathcal{L}aw}{\to} N\left(B(z), \alpha f^2(z)\right) \tag{41}$$

with

$$\lim_{n\to\infty} \operatorname{cov}(A_n, B_n) = 0. \tag{42}$$

Hence, the result will follow from equations (40), (41) and (42) applied to equation (39).

PROOF. (of equation (39)) In order to define A_n and B_n, first notice that by (K)-i

$$f_n(z) = \frac{cb}{n} \sum_{i=1}^{k} K\left(\frac{z - Y_i}{R(k)}\right) / \sum_{i=1}^{bk} \omega_i R^p(i)$$

where $Y_i = X_{i_j}$ with $i_j \in \{1, 2, \ldots, n\}$ for $j = 1, 2, \ldots, k$. Define

$$\begin{aligned} K_i &= K\left(\frac{z - Y_i}{R(k)}\right) \\ \tilde{E}(r) &= \mathrm{E}\left[K_i \,|R(k) = r\right] \\ \sigma^2(r) &= \operatorname{var}\left[K_i \,|R(k) = r\right]. \end{aligned}$$

Define also

$$Z_n = \sum_{i=1}^{k} \frac{K_i - \tilde{E}(R(k))}{\sqrt{k}\sigma(R(k))} \tag{43}$$

$$g_n^*(z) = (cbk/n) / \sum_{i=1}^{bk} \omega_i \lambda_i \tag{44}$$

Hence, A_n and B_n are given by:

$$A_n = \beta g_n^*(z)\sigma(R(k)) Z_n \tag{45}$$

$$B_n = \sqrt{k}\left\{\beta g_n^*(z)\tilde{E}(R(k)) - f(z)\right\} \tag{46}$$

□

PROOF. (of equation (40))
By corollary 2.1 we have (since $f_n = g_n^*$ when K is constant)

$$\text{as-}\lim_{n\to\infty} g_n^*(z) = f(z). \tag{47}$$

By lemma 2 in [8] we have as $n \to \infty$

$$Z_n \quad \stackrel{\mathcal{L}aw}{\to} \quad N(0,1) \tag{48}$$

$$\sigma^2(R(k)) \quad \stackrel{P}{\to} \quad \beta^{-1}\int K^2(y)\lambda(dy) - \beta^{-2} \tag{49}$$

provided that K is not constant on $S_0(1)$. Notice that if K is constant on $S_0(1)$ the asymptotic normality still holds. Thus, equation (40) follows from (47), (48), (49) together with Slutzky's theorem applied to (45). □

PROOF. (of equation (41)) This is the difficult part. We split it in two sub parts. Let us write

$$B_n \quad = \quad B_n^{(1)} + B_n^{(2)} \tag{50}$$

$$B_n^{(1)} \quad = \quad g_n^*(z)\sqrt{k}\left\{\beta\tilde{E}(R(k)) - 1\right\} \tag{51}$$

$$B_n^{(2)} \quad = \quad \sqrt{k}\left\{g_n^*(z) - f(z)\right\}. \tag{52}$$

We first show that

$$B_n^{(1)} \stackrel{P}{\to} (\text{BIAS1}). \tag{53}$$

Where (BIAS1) is defined by equation (61) bellow. We then show that

$$B_n^{(2)} \stackrel{\mathcal{L}aw}{\to} N\left((\text{BIAS2}), \alpha f^2(z)\right) \tag{54}$$

where (BIAS2) is defined by equation (73) bellow. Hence, (41) follows from Slutzky's theorem applied to (50) and we obtain $B(z) =$(BIAS1)+(BIAS2).

PROOF. (of equation (53))
Recall that

$$\tilde{E}(r) = \mathrm{E}\left\{K\left(\frac{z - Y_i}{r}\right)|R(k) = r\right\}$$

thus,

$$\begin{aligned}\tilde{E}(r) &= \int_{S(r)} K\left(\frac{z-y}{r}\right)\frac{f(y)\lambda(dy)}{H(r)} \\ &= \int_{S_0(1)} K(x)\frac{f(z-rx)}{H(r)}r^p\lambda(dx).\end{aligned} \tag{55}$$

By doing a second order Taylor expansion of $f(z-rx)$ about z and replacing in (55) we obtain,

$$\begin{aligned}\tilde{E}(r) &= \frac{f(z)}{H(r)}r^p + \\ & \frac{r^{p+2}}{2H(r)}\int_{S_0(1)} K(x)x'H_2(z)x\lambda(dx) + \frac{o\left(r^{p+2}\right)}{H(r)} \qquad (56)\end{aligned}$$

where we have used the hypotheses (K)-ii and the continuity of the Hessian $H_2(z)$. Expanding again $f(x)$ about z we can write

$$\begin{aligned}f(x) &= f(z) + (x-z)'\nabla f(z) + \frac{1}{2}(x-z)'H_2(z)(x-z) \\ & + \frac{1}{2}(x-z)'\left\{H_2(x^*) - H_2(z)\right\}(x-z) \qquad (57)\end{aligned}$$

where, as before, $H_2(x)$ denotes the Hessian of f at x and $x^* = z + \tau(x-z)$ for some $\tau \in [0,1]$. The assumed continuity of H_2 at z implies that $H_2(x^*) = H_2(z) + o(1)$ as $||x-z|| \to 0$. Thus, from (57) we have as $r \to 0$,

$$\begin{aligned}H(r) &= \int_{S(r)} f(x)\lambda(dx) \\ &= \beta f(z)r^p + \frac{1}{2}\gamma(z)r^{p+2} + o\left(r^{p+2}\right) \qquad (58)\end{aligned}$$

where we have used the hypotheses (N) on the norm and

$$\gamma(z) = \int_{S_0(1)} y'H_2(z)y\lambda(dy). \qquad (59)$$

Thus, from (56) and (58) we have

$$\begin{aligned}\tilde{E}(r) &= \frac{f(z)r^p + \frac{1}{2}\bar{\eta}(z)r^{p+2} + o\left(r^{p+2}\right)}{\beta f(z)r^p\left\{1 + \frac{1}{2}\bar{\gamma}(z)\beta^{-1}r^2 + o\left(r^2\right)\right\}} \\ &= \beta^{-1}\left\{1 + \frac{1}{2}\bar{\eta}(z)r^2 + o\left(r^2\right)\right\}\left\{1 - \frac{1}{2}\bar{\eta}(z)\beta^{-1}r^2 + o\left(r^2\right)\right\}.\end{aligned}$$

Hence,

$$\beta\tilde{E}(r) = 1 + \frac{1}{2}\left\{\bar{\eta}(z) - \beta^{-1}\bar{\gamma}(z)\right\}r^2 - \frac{1}{4}\bar{\eta}(z)\bar{\gamma}(z)\beta^{-1}r^4 + o\left(r^4\right) \qquad (60)$$

where $\bar{\gamma}(z) = \gamma(z)/f(z)$ and

$$\bar{\eta}(z) = \int_{S_0(1)} K(x)x'H_2(z)x\lambda(dx)/f(z).$$

From the consistency of the k-nn (see lemma 2.2), equations (47), (60) and the hypothesis (28) of the theorem we obtain (53) with $\eta(z)$ given by (31) and

$$(\text{BIAS1}) = f(z)\frac{1}{2}\beta\eta(z)\left[\beta f(z)\right]^{-1-2/p} a^{\frac{4+p}{2p}}. \tag{61}$$

□

PROOF. (of equation (54))
Notice that from (44)

$$\begin{aligned} g_n^*(z) &= \frac{cbk/n}{\sum_{i=1}^{bk}\omega_i H(R(i))}\cdot\frac{\sum_{i=1}^{bk}\omega_i H(R(i))}{\sum_{i=1}^{bk}\omega_i\lambda_i} \\ &\equiv V_n^*(z)\cdot W_n^*(z) \end{aligned} \tag{62}$$

we can write

$$\sum_{i=1}^{bk}\omega_i H(R(i)) = \sum_{i=1}^{bk}\left(\sum_{j=1}^{bk}\omega_j\right) D_i$$

where $D_i = H(R(i)) - H(R(i-1))$ with $R(0)\equiv 0, R(n+1) = +\infty$. It is well known that the joint distribution of the sample coverages $D_1, D_2, \ldots, D_{n+1}$ is the same as the joint distribution of $Q_1/S_{n+1}, Q_2/S_{n+1}, \ldots, Q_{n+1}/S_{n+1}$; where the Q_i's are iid exponentials with mean one and $S_{n+1} = \sum_{i=1}^{n+1} Q_i$. Therefore,

$$\begin{aligned} V_n^*(z) &\stackrel{d}{=} \frac{cbk/n}{\sum_{i=1}^{bk}\left(\sum_{j=1}^{bk}\omega_j\right)\frac{Q_i}{S_{n+1}}} \\ &= \frac{c}{\frac{1}{bk}\sum_{i=1}^{bk} i\omega_i}\left[\sum_{j=1}^{bk} c_j^* Q_j\right]^{-1}\left(\frac{S_{n+1}}{n}\right) \end{aligned}$$

where

$$c_j^* = \sum_{j=1}^{bk}\omega_j/\sum_{i=1}^{bk} i\omega_i. \tag{63}$$

Let us call $\psi_n^* = \sum_{j=1}^{bk} c_j^* Q_j$, then

$$V_n^*(z) \stackrel{d}{=} \frac{S_{n+1}/n}{\psi_n^*} t_n$$

with $t_n = cbk / \sum_{i=1}^{bk} i\omega_i$. Thus,

$$\begin{aligned} \sqrt{k}\left(V_n^*(z) - 1\right)\psi_n^* &\stackrel{d}{=} t_n \sqrt{\frac{k}{n}}\sqrt{n}\left(\frac{S_{n+1}}{n} - 1\right) \\ &\quad - t_n\sqrt{k}\left(\psi_n^* - 1\right) + \psi_n^*\sqrt{k}(t_n - 1). \end{aligned} \tag{64}$$

Notice that

$$\lim_{n\to\infty} \frac{1}{bk}\sum_{i=1}^{bk} i\omega_i = \int_0^1 u\omega(u)du = c \tag{65}$$

from where $t_n \to 1$ as $n \to \infty$ and this together with the CLT applied to S_{n+1} and the fact that $k/n \to 0$ imply that the first term of (64) goes to zero in probability as $n \to \infty$. The last term in (64) also approaches zero in probability since

$$\begin{aligned} \frac{1}{bk}\sum_{i=1}^{bk} i\omega_i - c &= \sum_{i=1}^{bk}\int_{(i-1)/bk}^{i/bk}\left(\frac{i}{bk} - u\right)\omega(u)du \\ &\leq \frac{1}{bk}\int_0^1 \omega(u)du = \frac{1}{bk}. \end{aligned} \tag{66}$$

The asymptotic normality of the second term in (64) follows from Slutzky's theorem. We have that $bkc_j^* \leq c^{-1}$ for all n (the left hand side of (66) is always positive). Thus, for all $\epsilon > 0$

$$\sqrt{k}\sum_{j=1}^{bk} \mathrm{P}\left[c_j^* Q_j \geq \epsilon\right] = \sqrt{k}\sum_{j=1}^{bk} e^{-\epsilon/c_j^*} \leq \sqrt{k}e^{-\epsilon cbk}$$

which goes to zero as $n \to \infty$ which implies that ψ_n^* is asymptotically normal (see [5] p.316). Applying again Slutzky's theorem to (64) we obtain the following lemma:

Lemma 3.1 *If $k(n) \to \infty$ and $k/n \to 0$ then as $n \to \infty$*

$$\sqrt{k}\left(V_n^*(z) - 1\right) \stackrel{\mathcal{L}aw}{\to} N(0, \alpha)$$

where

$$\alpha = \left(cb^2\right)^{-1}\int_0^1 \left\{\int_u^1 \omega(t)dt\right\}^2 du \tag{67}$$

PROOF. (of equation (67))

$$\alpha = \lim_{n\to\infty} k\mathrm{var}\left(V_n^*(z)\right) = \lim_{n\to\infty} k\mathrm{var}\left(\psi_n^*\right) = \lim_{n\to\infty} k\sum_{j=1}^{bk}\left(c_j^*\right)^2$$

the expression for α follows by replacing (63) and (65) above. □

Notice that from (52) and (62) we can write

$$B_n^{(2)} = \sqrt{k} W_n^*(z) \left(V_n^*(z) - 1 \right) + \sqrt{k} \left(W_n^*(z) - f(z) \right). \tag{68}$$

Hence, (54) follows from lemma 3.1 (above) and lemma 3.2 (below) and Slutzky's theorem applied to (68). □

Lemma 3.2

$$\sqrt{k} \left(W_n^*(z) - f(z) \right) \xrightarrow{P} \textit{(BIAS2)}$$

PROOF. By (58) we obtain

$$W_n^*(z) = f(z) + \left[\frac{\sum_{i=1}^{bk} \omega_i \lambda_i R^2(i)}{\sum_{i=1}^{bk} \omega_i \lambda_i} \right] \frac{\gamma(z)}{2\beta} + o_p \left(R^2(bk) \right). \tag{69}$$

Where the little o_p (above) means *in probability*. Notice that

$$\begin{aligned} \sqrt{k} R^2(bk) &= \left[\frac{bk/n}{R^p(bk)} \right]^{-2/p} \left(\frac{bk}{n} \right)^{-2/p} \\ &\xrightarrow{P} \frac{a^{\frac{p+4}{2}}}{[\beta f(z)]^{2/p}} b^{2/p} \end{aligned}$$

where we have used the hypothesis (28) of the theorem. Therefore,

$$\sqrt{k} o_p \left(R^2(bk) \right) = o_p \left(n^0 \right). \tag{70}$$

We also have that

$$T_n = \sqrt{k} \frac{\sum_{i=1}^{bk} \omega_i \lambda_i R^2(i)}{\sum_{i=1}^{bk} \omega_i \lambda_i} \leq \sqrt{k} R^2(bk).$$

Thus, T_n is bounded in probability but the limit will depend on the specific weight function $\omega(u)$ used. Let $N > 0$ be a given integer such that $[bk/N] = m \geq 1$. From (N)-i and by the consistency of the k-nn (see lemma 2.2) we can write for $i = m, m+1, \ldots, bk$

$$R(i) = (\beta f(z))^{-1/p} \left(\frac{i}{n} \right)^{1/p} + o_p \left(\left(\frac{i}{n} \right)^{1/p} \right) \tag{71}$$

hence,

$$T_n = \left\{ \frac{cbk/n}{\sum_{i=1}^{bk} \omega_i \lambda_i} \right\} c^{-1}\sqrt{k}\left(\frac{bk}{n\beta}\right)^{2/p} \sum_{i=m}^{bk} \omega_i \left[\frac{i/n}{\beta R^p(i)}\right]^{-1-2/p} \left(\frac{i}{bk}\right)^{1+2/p}$$

by corollary 2.1 (with uniform kernel K) the expression in curly brackets equals $f(z) + o_p(n^0)$. Thus, after splitting the sum, replacing equation (71), and regrouping we have:

$$T_n = [\beta f(z)]^{-2/p}\sqrt{k}\left(\frac{bk}{n}\right)^{2/p} c^{-1}\sum_{i=1}^{bk}\omega_i\left(\frac{i}{bk}\right)^{1+2/p} + I_N + o_p(n^0) \qquad (72)$$

where

$$I_N \le O_p\left(n^0\right)\int_0^{1/N} u^{1+2/p}\omega(u)du + o_p(n^0)$$

thus, if ω has at most a jump discontinuity at $u = 0$, we have

$$I_N = O_p(n^0)o\left(N^0\right) + o_p(n^0).$$

From here, (72), and the hypothesis of the theorem (28), we obtain

$$T_n = [\beta f(z)]^{-2/p} b^{2/p} a^{\frac{p+4}{2p}} c^{-1}\int_{1/N}^{1} u^{1+2/p}\omega(u)du + O_p(n^0)o(N^0) + o_p(n^0).$$

By first letting $n \to \infty$ and then $N \to \infty$ we have

$$T_n \xrightarrow{P} [\beta f(z)]^{-2/p} b^{2/p} a^{\frac{p+4}{2p}} c^{-1}\int_0^1 u^{1+2/p}\omega(u)du.$$

From this, (70), and (69) we obtain lemma 3.2 with

$$(\text{BIAS2}) = \left\{\frac{\gamma(z)}{2\beta}\right\}[\beta f(z)]^{-2/p} b^{2/p} a^{\frac{p+4}{2p}} c^{-1}\delta. \qquad (73)$$

This completes (54) □

PROOF. (of equation (42)) Notice that by the definitions of A_n and B_n (see (45) and (46)) the consistency of $g_n^*(z)$ (see (47)) and the convergence of $\sigma(R(k))$ (see (49)) we only need to show that Z_n is asymptotically uncorrelated to $R(k)$, but

$$\begin{aligned} \operatorname{cov}(Z_n, R(k)) &= \mathrm{E}\left(Z_n R(k)\right) \\ &= \mathrm{E}\left\{R(k)\mathrm{E}\left[Z_n | R(k)\right]\right\} \\ &= \mathrm{E}\left\{R(k)\cdot 0\right\} = 0, \end{aligned}$$

holds not only asymptotically but for all n. □

PROOF. (of theorem 3.2)
The proof of theorem 3.1 will be valid for this version provided that we rewrite (41). In order to show (41) it is enough to consider the second order Taylor expansion of $f(x)$ and $f(z-rx)$ about z in the form

$$f(x) = \sum_{0\leq|j|\leq 2} \frac{f^{(j)}(z)}{j!}(x-z)^j + \sum_{|j|=2} \frac{(x-z)^j}{j!}\left[f^{(j)}(x^*) - f^{(j)}(z)\right]$$

where $||x^* - z|| \leq ||x - z||$ and,

$$f(z-rx) = \sum_{0\leq|j|\leq 2} (-1)^{|j|} r^{|j|} \frac{j^{(j)}(z)}{j!} x^j + r^2 \sum_{|j|=2} \frac{x^j}{j!}\left[f^{(j)}(y^*) - f^{(j)}(z)\right]$$

where $||x^* - z|| \leq r||x||$. This allows the application of the hypothesis given by equation (35). □

4 Optimal ω and Asymptotic Mean Square Error Comparisons

Optimizing the weight function ω

General optimization of the weight function ω is in principle possible from the expressions for the asymptotic variance and bias given in theorem 3.1. The general solution seems (at present to the author) to be a difficult problem. However, the following special case is tractable:

Consider an estimator of the type (10). We shall try to find weights $\omega_1, \omega_2, \ldots, \omega_{bk}$ such that the asymptotic variance of the estimator is minimized. By theorem 3.1 (see the proof of equation (67)) this is equivalent to find the $\omega_i' s$ such that

$$\alpha = \lim_{n\to\infty} \sum_{j=1}^{bk} \left\{ \sum_{i=j}^{bk} \omega_i / \sum_{i=1}^{bk} i\omega_i \right\}^2$$

is minimum. This problem turns out to have a very simple solution if we restrict the search of ω_i's to those such that

$$\sum_{i=1}^{bk} i\omega_i = cbk.$$

(Asymptotically this is equivalent to find the weight function ω that minimizes α among all ω's with fixed mean c). Therefore, we look for the solution to the following problem P).

$$P)\quad \text{Find the } \omega_i\text{'s that minimize} \sum_{j=1}^{bk}\left(\sum_{i=j}^{bk}\omega_i\right)^2$$

subject to the constraints:

i. $\sum_{i=1}^{bk}\omega_i = 1$

ii. $\sum_{i=1}^{bk} i\omega_i = cbk.$

Therefore to solve problem P) is equivalent to find the minimum of the Lagrangian $L = L(\omega_1, \dots, \omega_{bk}, \alpha_1, \alpha_2)$ given by

$$L = \sum_{j=1}^{bk}\left(\sum_{i=j}^{bk}\omega_i\right)^2 - 2\alpha_1\left(\sum_{i=1}^{bk}\omega_i - 1\right) - 2\alpha_2\left(\sum_{i=1}^{bk} i\omega_i - cbk\right) \tag{74}$$

where α_1 and α_2 are the Lagrange multipliers associated to the constraints. The optimality conditions are given by

$$\frac{\partial L}{\partial \omega_r} = \sum_{j=1}^{r} 2\sum_{i=j}^{bk}\omega_i - 2\alpha_1 - 2r\alpha_2 = 0, \quad \text{for } r = 1, 2, \dots, bk \tag{75}$$

$$\frac{\partial L}{\partial \alpha_1} = \frac{\partial L}{\partial \alpha_2} = 0 \quad \text{(the constraints)}$$

interchanging the order of summation we find

$$\sum_{j=1}^{r}\sum_{i=j}^{bk}\omega_i = \sum_{i=1}^{r} i\omega_i + \sum_{i=r+1}^{bk} r\omega_i = \sum_{i=1}^{r} i\omega_i + r\left\{1 - \sum_{i=1}^{r}\omega_i\right\}$$

hence,

$$\sum_{j=1}^{r}\sum_{i=j}^{bk}\omega_i = r + \sum_{i=1}^{r}(i-r)\omega_i. \tag{76}$$

Therefore, from (75) and (76) the optimal weights must satisfy

$$\sum_{i=1}^{r}(r-i)\omega_i = r - (\alpha_1 + r\alpha_2) \quad \text{for } r = 1, 2, \dots, bk. \tag{77}$$

From (77) when $r = 1$ we get $\alpha_1 + \alpha_2 = 1$, hence, using this and (77) we can write

$$\omega_j = j(1-\alpha_2) - \sum_{i=1}^{j-1}(j-i+1)\omega_i \quad \text{for } j = 2, 3, \ldots, bk-1. \tag{78}$$

From (77) when $r = 2$ we obtain $\omega_1 = 1 - \alpha_2$, hence using (78) we have $\omega_2 = 0$ and

$$\omega_j = -\sum_{i=2}^{j-1}(j-i+1)\omega_i \quad \text{for } j = 3, 4, \ldots, bk-1$$

and these equations imply $\omega_j = 0, \forall j = 2, 3, \ldots, bk-1$. Hence, from (77) with $r = bk$ we obtain $\omega_{bk} = \alpha_2$ which is in turn determined from the constraints. Therefore, the optimal weights are given by

$$\omega_1 = \frac{bk(1-c)}{bk-1}, \quad \omega_2 = \ldots = \omega_{bk-1} = 0, \quad \omega_{bk} = \frac{cbk-1}{bk-1}. \tag{79}$$

Replacing (79) into the expression for α we find the optimum $\alpha = \alpha^*$ given by

$$\alpha^* = \lim_{k\rightarrow} \frac{k}{(cbk)^2}\left(1 + \frac{(cbk-1)^2}{bk-1}\right) = b^{-1}. \tag{80}$$

Asymptotic mean square error comparisons

Theorem 3.1 can be used to show that the class of estimators (10) dominates the generalized k-nn class (see [7]) and thus, the classical kernel as well. To see this choose $k = an^{\frac{4}{p+4}}$. From theorem 3.1 we can write as $n \rightarrow \infty$

$$\mathrm{var}\left[f_n(z)\right] = k^{-1}\left\{\beta f^2(z)\int K^2(x)\lambda(dx) - (1-\alpha)f^2(x)\right\} + o\left(k^{-1}\right). \tag{81}$$

We also have from theorem 3.1 that as $n \rightarrow \infty$

$$\mathrm{E}\left[f_n(z)\right] = f(z) + k^{-1/2}B(z) + o\left(k^{-1/2}\right). \tag{82}$$

Replacing the expression for $B(z)$ and using the notation in [7], $Q(f)(z) = \eta(z) + \beta^{-1}\gamma(z)$ we have

$$E\left[f_n(z)\right] = f(z) + \frac{Q(f)(z)}{2[\beta f(z)]^{2/p}}\left(\frac{k}{n}\right)^{2/p} - q\frac{\gamma(z)}{2[\beta f(z)]^{2/p}}\left(\frac{k}{n}\right)^{2/p} + o\left(\left(\frac{k}{n}\right)^{2/p}\right) \tag{83}$$

with

$$q = \left(1 - b^{2/p}\frac{\delta}{c}\right)\beta^{-1}. \tag{84}$$

Therefore, when $b = c = \delta = 1$ (i.e. no weighting in the denominators) we recover the results in [7] as a special case of equations (81) and (83). Moreover, if we choose

$$b = \left(\frac{c}{\delta}\right)^{p/2} \tag{85}$$

(which is a proper choice since $b \geq 1$) we have from (84) that $q = 0$ and the asymptotic expression for the bias of the generalized k-nn coincides with (83).

A global measure of the asymptotic relative efficiency of our estimators with respect to the generalized k-nn can be obtained by computing the ratio of the asymptotic mean square error expression (see [10]) $(\text{AMSE})_N, (\text{AMSE})_\omega$ for the generalized k-nn and ours respectively. We obtain,

$$\begin{aligned}
(\text{AMSE})_N n^{\frac{4}{p+4}} &= \left\{\beta f^2(z) \int K^2(x)\lambda(dx)\right\} a^{\frac{-4}{p+4}} + \left\{\frac{Q^2(f)(z)}{4[\beta f(z)]^{4/p}}\right\} a^{\frac{16}{p(p+4)}} \\
&\equiv A_N a^{\frac{-4}{p+4}} + B_N a^{16p(p+4)} \qquad (86)\\
(\text{AMSE})_\omega n^{\frac{4}{p+4}} &= \left\{\beta f^2(z) \int K^2(x)\lambda(dx) - (1-\alpha) f^2(z)\right\} a^{\frac{-4}{p+4}} \\
&\quad + \left\{\frac{Q(f)(z)}{2[\beta f(z)]^{2/p}} - q\frac{\gamma(z)}{2[\beta f(z)]^{2/p}}\right\} a^{\frac{16}{p(p+4)}} \\
&\equiv A_\omega a^{\frac{-4}{p+4}} + B_\omega a^{\frac{16}{p(p+4)}}. \qquad (87)
\end{aligned}$$

From (86) and (87) we obtain that the a that minimizes the (AMSE) has the form

$$a^{\frac{4}{p}} = \frac{pA}{4B}. \tag{88}$$

Replacing the optimal a's in (86) and (87) we obtain the $(\text{AMSE})^*$ expressions satisfying

$$\frac{(\text{AMSE})_N^*}{(\text{AMSE})_\omega^*} = \left\{\left(\frac{A_N}{A_\omega}\right)^4 \left(\frac{B_N}{B_\omega}\right)^p\right\}^{\frac{1}{p+4}}$$

or equivalently,

$$\frac{(\text{AMSE})_N^*}{(\text{AMSE})_\omega^*} = \left\{\left(\frac{\beta\int K^2(x)\lambda(dx)}{\beta\int K^2(x)\lambda(dx) - (1-\alpha)}\right)^4 \left(\frac{Q(f)(z)}{Q(f)(z) - q\gamma(z)}\right)^{2p}\right\}^{\frac{1}{p+4}} \tag{89}$$

and if we choose b satisfying (85) (i.e. $q = 0$) we obtain:

$$(\text{ARE}) = \frac{(\text{AMSE})_N^*}{(\text{AMSE})_\omega^*} > \left(1 + \frac{1-\alpha}{\beta\int K^2(x)\lambda(dx)}\right)^{\frac{4}{p+4}} > 1 \tag{90}$$

and the above inequality holds uniformly in $f \in \mathcal{C}^2, z \in \mathbb{R}^p, \beta, K, c, p$, and ω. In other words (90) shows that no matter what smoothness parameter is used in the generalized k-nn estimator with kernel K an estimator from the class (10) with b as in (85) and the same kernel, K, has a smaller (AMSE) for some values of k. Even though the choice of $q = 0$ allows to show that there is improvement over the classical methods, the lower bound is not best possible (in fact it is always less than two). For example, consider the case of uniform kernel and uniform weight function ω; i.e. $\delta/c = p/(p+1), \alpha = 1/3, \beta \int K^2 = 1, b = (1 + 1/p)^{p/2}$ and, ARE $= (1.5)^{4/(p+4)}$ that produces the sequence: $1.38, 1.31, 1.26 \ldots, 1$ for $p = 1, 2, 3, \ldots, \infty$. However (89) shows that in principle the lower bound can be increased without limit by choosing $q = Q(f)(z)/\gamma(z)$ adaptively at each point z.

References

[1] L. Devroye. *A course in density estimation.* Birkhäuser, Boston, 1987.

[2] L. Devroye and L. Györfi. *Nonparametric Density Estimation: The L_1 View.* John Wiley and Sons, 1985.

[3] E. Fix and J.L. Hodges. Discrimination analysis. nonparametric discrimination: Consistency properties. Report 4, Project Number 21-49-004, USAF school of Aviation medicine, Randolf Field, Texas, 1951.

[4] J. Kiefer. Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. In *Proc. sixth Berkeley Symp. Math. Stat. Prob.*, volume 1, pages 227–244, 1972.

[5] M. Loève. *Probability theory.* Van Nostrand, Princeton, 1960.

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

[7] Y.P. Mack and M. Rosenblatt. Multivariate k-nearest neighbor density estimates. *Journ. Mult. Anal.*, 9:1–15, 1979.

[8] D.S. Moore and J.W. Yackel. Large sample properties of nearest neighbor density function estimators. In S.S. Gupta and D.S. Moore, editors, *Statistical decision theory and related topics.* Academic Press, 1976.

[9] C. Rodríguez and J. Van Ryzin. Maximum entropy histograms. *Stat. and Prob. letters*, 3:117–120, 1985.

[10] C. Rodríguez and J. Van Ryzin. Large sample properties of maximum entropy histograms. *IEEE Trans. Inf. thr.*, IT-32 No.6:751–759, 1986.