# ASYMPTOTIC PROPERTIES OF KERNEL ESTIMATES OF A REGRESSION FUNCTION

Włodzimierz GREBLICKI and Adam KRZYŻAK

Technical University of Wrocław,
Institute of Engineering Cybernetics, Wrocław, Poland

printed in:
Journal of Statistical Planning and Inference
Vol. 4, No. 1, pp. 81–90, 1980

*Abstract:* The consistency of the Parzen kernel type as well as a recursive kernel estimates of a regression function is shown. The rates of the convergence are studied and compared. Moreover, the problem of selecting asymptotically kernels is discussed.

*Key words and phrases:* Nonparametric Estimate; Estimate; Kernel-type Estimate; Regression Function; Consistency

## 1. Introduction

Let $(X, Y)$ be a pair of random variables valued in $R^d$ and $R$, respectively and let $g$ be the marginal density of $X$. Given a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of independent observations of $(X, Y)$ we estimate the regression of $Y$ on $X$. We consider the following four estimates:

$$\hat{R}_n(x) = \frac{1}{nh^p(n)g(x)} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h(n)}\right), \tag{1}$$

$$\tilde{R}_n(x) = \frac{1}{ng(x)} \sum_{i=1}^{n} \frac{1}{h^p(i)} Y_i K\left(\frac{x - X_i}{h(i)}\right), \tag{2}$$

if $g$ is known, and

$$\hat{S}_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h(n)}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h(n)}\right)}, \tag{3}$$

$$\tilde{S}_n(x) = \frac{\sum_{i=1}^{n} \frac{1}{h^p(i)} Y_i K\left(\frac{x - X_i}{h(i)}\right)}{\sum_{i=1}^{n} \frac{1}{h^p(i)} K\left(\frac{x - X_i}{h(i)}\right)}, \tag{4}$$

if $g$ is unknown, where the kernel $K$ and the sequence $\{h_n\}$ are selected by a statistician.

Estimate (3) has been studied by Watson (1964), Nadaraya (1964, 1970) and Noda (1976), while its recursive version (4) has been examined by Ahmad and Lin (1976).

A Borel measurable kernel satisfies the following conditions:

$$\int K(x)dx = 1, \tag{5}$$

$$\sup_x |K(x)| < \infty, \int |K(x)|dx < \infty, \lim_{\|x\|\to\infty} \|x\|^p \, |K(x)| = 0. \tag{6}$$

On the sequence of positive numbers the following conditions will be imposed:

$$h(n) \to 0, \tag{7}$$

$$nh^p(n) \to \infty, \tag{8}$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^p(i)} \to 0, \text{ as } n \to \infty. \tag{9}$$

## 2. Preliminaries

In the paper $b$ and var denote bias and variance of a random variable, respectively. For a vector $x$, $x_i$ is its $ith$ coordinate. $C(f_1, \ldots, f_m)$ is a set of all points of continuity of all functions $f_1, \ldots, f_m$; $a_n \sim b_n$ denotes that $a_n/b_n \to 1$ as $n \to \infty$.

First of all we shall give a lemma which is based on Parzen's (1962) Theorem 1A and Cacoullos' Theorem 2.1.

Denote

$$R_s(x) = E\{Y^s/X = x\},$$

and define $R_{sn}(x)$ by the following:

$$g(x)R_{sn}(x) = \frac{1}{h^p(n)} \int R_s(x-z)g(x-z)K\left(\frac{z}{h(n)}\right) dz.$$

**Lemma 1.** *If $E|Y|^s < \infty$, (6) and (7) are satisfied, then*

$$\lim_{n\to\infty} R_{sn}(x) = R_s(x) \int K(z)dz$$

*at every point $x \in C(g, R_s)$.*

**Proof.** Notice that

$$\begin{aligned} I_n &= g(x)\left[R_{sn}(x) - R_s(x)\int K(z)dz\right] \\ &= \frac{1}{h^p(n)} \int \left[g(x-z)R_s(x-z) - g(x)R_s(x)\right] K\left(\frac{z}{h(n)}\right) dz. \end{aligned}$$

Choose $\delta > 0$ and split the region of integration into two sets $\|z\| \le \delta$, $\|z\| > \delta$. Then

$$\begin{aligned}|I_n| &\le \sup_{\|x\|\le\delta} |g(x-z)R_s(x-z) - g(x)R_s(x)| \int |K(z)|dz \\ &+ \int_{\|x\|>\delta} \frac{1}{\|z\|^p} \frac{\|z\|^p}{h^p(n)} \left| K\left(\frac{z}{h(n)}\right)\right| |g(x-z)R_s(x-z)|dz \\ &\le \sup_{\|x\|\le\delta} |g(x-z)R_s(x-z) - g(x)R_s(x)| \int |K(z)|dz \\ &+ \frac{1}{\delta^p} E|Y|^s \sup_{\|x\|>\delta/h(n)} \|z\|^p |K(z)| + |g(x)R_s(x)| \int_{\|x\|>\delta/h(n)} |K(z)|dz.\end{aligned}$$

It is clear that for every $\varepsilon > 0$ there exists $h > 0$ such that the above is not greater than $\varepsilon$, if $h(n) < h$. Thus the proof is complete.

**Corollary 1**. *If $E|Y| < \infty$, then, under (6) and (7),*

$$\lim_{n\to\infty} \frac{1}{h^p(n)} E\left[YK\left(\frac{x-X}{h(n)}\right)\right] = g(x)R(x)\int K(z)dz$$

*at every $x \in C(g,R)$.*

Denote

$$\operatorname{var}[Y|X=x] = \sigma^2(x).$$

**Corollary 2**. *If $EY^2 < \infty$, and (6), (7) are satisfied, then*

$$\lim_{n\to\infty} \frac{1}{h^p(n)} \operatorname{var}\left[YK\left(\frac{x-X}{h(n)}\right)\right] = g(x)\left[R^2(x) + \sigma^2(x)\right]\int K^2(z)dz$$

*at every $x \in C(g,R^2,\sigma^2)$.*

## 3. The consistency of the estimate

In this section we establish the pointwise consistency of all the estimates. On the contrary to Noda (1976) and Ahmad and Lin (1976) we do not assume the conditional density of $Y$ given $X = x$ to exist. We examine asymptotic properties of the estimates at a point $x$ at which $g(x) > 0$.

It is quite easy to show the consistency of $\hat{R}_n(x)$, since by Corollary 1, we get:

**Lemma 2**. (asymptotic unbiasedness of $\hat{R}_n(x)$). *If $E|Y| < \infty$, and (5), (6) and (7) hold, then*

$$\lim_{n\to\infty} E\hat{R}_n(x) = R(x)$$

*at every $x \in C(g,R)$.*

By Corollary 2, we have:

**Lemma 3.** (variance of $\hat{R}_n(x)$). *If $EY^2 < \infty$, (6), (7) and (8) are satisfied, then*

$$\lim_{n\to\infty} nh^p(n)\text{var}\left[\hat{R}_n(x)\right] = \frac{R^2(x)+\sigma^2(x)}{g(x)}\int K^2(z)dz$$

*at every $x \in C(g, R^2, \sigma)$.*

Lemmas 2 and 3 imply

**Theorem 1.** (mean square consistency of $\hat{R}_n(x)$). *If $EY^2 < \infty$ and (5)–(8) hold, then*

$$\lim_{n\to\infty} E\left[\hat{R}_n(x) - R(x)\right]^2 = 0$$

*at every $x \in C(g, R, \sigma)$.*

**Theorem 2.** (consistency of $\hat{S}_n(x)$). *Under the assumptions of Theorem 1*

$$\hat{S}_n(x) \to R(x) \textit{ in probability}$$

*as $n \to \infty$, at every $x \in C(g, R, \sigma)$.*

**Proof.** Cacoullos (1966, Theorem 3.2) showed that

$$\frac{1}{nh^p(n)}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h(n)}\right) \to g(x) \text{ in probability}$$

as $n \to \infty$, at every $x \in C(g)$. The above and Theorem 1 imply the consistence of $\hat{S}_n(x)$.

**Remark.** Noda's (1976) proof of the mean square consistency of $\hat{S}_n(x)$ seems not to be correct, since the inequality in the proof of Theorem 3.1 is not right.

We shall now investigate recursive estimates (2) and (4). The next lemma can be easily proved by Corollary 1.

**Lemma 4.** (asymptotic unbiasedness of $\tilde{R}_n(x)$). *If $E|X| < \infty$, then under (5)–(7),*

$$\lim_{n\to\infty} E\tilde{R}_n(x) = R(x)$$

*at every $x \in C(g, R)$.*

**Lemma 5.** (variance of $\tilde{R}_n(x)$). *If $EY^2 < \infty$, (6), (7) and (9) hold, then*

$$\lim_{n\to\infty} \frac{n^2}{\sum_{i=1}^n h^{-p}(i)}\text{var}[\tilde{R}_n(x)] = \frac{R^2(x)+\sigma^2(x)}{g(x)}\int K^2(z)dz$$

*at every $x \in C(g, R^2, \sigma)$.*

**Proof.** Obviously

$$\frac{1}{\sum_{i=1}^n h^{-p}(i)} g^2(x)\text{var}[\tilde{R}_n(x)] = \frac{\sum_{i=1}^n h^{-p}(i)\left\{h^{-p}(i)\text{var}\left[YK\left(\frac{x-X}{h(i)}\right)\right]\right\}}{\sum_{i=1}^n h^{-p}(i)}$$

tends to the same limit as

$$\frac{1}{h^p(i)}\text{var}\left[YK\left(\frac{x-X}{h(i)}\right)\right]$$

since $\sum_{i=1}^{n} h^{-p}(i) \to \infty$ as $n \to \infty$. This and Corollary 2 complete the proof.

Lemmas 4 and 5 yield:

**Theorem 3.** (mean square consistency of $\tilde{R}_n(x)$). *If* $EY^2 < \infty$*, (5), (6), (7) and (9) hold, then*

$$\lim_{n\to\infty} E\left[\tilde{R}_n(x) - R(x)\right]^2 = 0$$

*at every* $x \in C(g, R, \sigma)$.

Applying Cacoullos's (1966) Theorem 2.1 one can show

**Theorem 4.** (consistency of $\tilde{S}_n(x)$). *Under all the assumptions of Theorem 3*

$$\tilde{S}_n(x) \to R(x) \textit{in probability}$$

*at every* $x \in C(g, R, \sigma)$.

## 4. The rate of convergence of $\hat{R}_n(x)$ and $\tilde{R}_n(x)$

For simplicity suppose that

$$K(x_1, \ldots, x_p) = \prod_{i=1}^{p} H(x_i)$$

and moreover

$$H(v) = H(-v), \int v^6 H(v) dv < \infty, \tag{10}$$

and both $g$ and $R$ have partial derivatives up to the third order in a closed neighborhood of a point $x$ at which both $g$ and $R$ are continuous and $g(x) > 0$. Since

$$b\left[\frac{1}{h^p(n)} YK\left(\frac{x-X}{h(n)}\right)\right] = \int \left[g(x - h(n)z)R(x - h(n)z) - g(x)R(x)\right] K(z) dz$$

then by expanding both $g(x-h(n)z)$ and $R(x-h(n)z)$ by Taylor's theorem and using (10), one gets

$$b\left[\frac{1}{h^p(n)} YK\left(\frac{x-X}{h(n)}\right)\right] = \sqrt{c_1} h^2(n) + o(h^2(n)), \tag{11}$$

where

$$\sqrt{c_1} = \left[\sum_{i=1}^{p} \frac{\partial g(x)}{\partial x_i}\frac{\partial R(x)}{\partial x_i} + \frac{R(x)}{2}\sum_{i=1}^{p}\frac{\partial^2 g(x)}{\partial x_i^2} + \frac{g(x)}{2}\sum_{i=1}^{p}\frac{\partial^2 R(x)}{\partial x_i^2}\right] \int v^2 H(v) dv.$$

Thus, by Lemma 3:

$$E\left[\hat{R}_n(x) - R_n(x)^2\right] \sim c_1 h^4(n) + c_2 n^{-1} h^{-p}(n),$$

where

$$c_2 = \frac{R^2(x) + \sigma^2(x)}{g(x)} \left( \int H^2(v) dv \right)^p .$$

Hence, for $h(n) = kn^{-\alpha}$, $0 < \alpha < 1/p$:

$$E\left[\hat{R}_n(x) - R_n(x)^2\right] \sim c_1 k^4 n^{-4\alpha} + \frac{c_2}{k^p} n^{-(1-\alpha p)}. \tag{12}$$

For $\alpha = \alpha^* = 1/(p+4)$, which minimizes the right side of (12),

$$E\left[\hat{R}_n(x) - R_n(x)^2\right] \sim \hat{c} n^{-4/(p+1)},$$

where

$$\hat{c} = c_1 k^4 + c_2 k^{-p}.$$

For the recursive estimate, by (11):

$$b\left[\tilde{R}_n(x)\right] = \sqrt{c_1} \frac{1}{n} \sum_{i=1}^{n} h^2(i) + \frac{1}{n} \sum_{i=1}^{n} o\left(h^2(i)\right).$$

Obviously for $h(n) = kn^{-\alpha}$, $0 < \alpha < 1/2$;

$$b\left[\tilde{R}_n(x)\right] \sim \sqrt{c_1} \frac{1}{n} \sum_{i=1}^{n} h^2(i) \sim \frac{\sqrt{c_1} k^2}{(1-2\alpha)} n^{-2\alpha}.$$

Hence, by Lemma 5, for $0 < \alpha < \min(1/2, 1/p)$;

$$E\left[\tilde{R}_n(x) - R_n(x)\right]^2 \sim \frac{c_1 k^4}{(1-2\alpha)^2} n^{-4\alpha} + \frac{c_2}{k^p (1+\alpha p)} n^{-(1-\alpha p)} \tag{13}$$

Finally, for $0 < \alpha < \min(1/2, 1/p)$;

$$\xi(\alpha) \backsim \frac{E\left[\tilde{R}_n(x) - R_n(x)\right]^2}{E\left[\hat{R}_n(x) - R_n(x)\right]^2}$$

where

$$\xi(\alpha) = \begin{cases} \dfrac{1}{1+\alpha p}, & \text{for } \alpha^* < \alpha < \min(1/2, 1/p), \\ \dfrac{\tilde{c}}{\hat{c}} & \text{for } \alpha = \alpha^*, \\ \dfrac{1}{(1-2\alpha)^2}, & \text{for } 0 < \alpha < \alpha^*, \end{cases}$$

and where

$$\tilde{c} = \frac{c_1 k^4 (p+4)^2}{(p+2)^2} + \frac{c_2 (p+4)}{k^p (2p+4)}.$$

Notice that $\xi(\alpha) < 1$ if $\alpha$ is greater that $\alpha^*$ and $\xi(\alpha) > 1$ for $\alpha$ less than $\alpha^*$. It is not possible to evaluate $\xi(\alpha^*)$ since neither $c_1$ nor $c_2$ are known.

The properties of the estimates depend also on the kernel $H$. Obviously a kernel $H_1$ is asymptotically not worse than $H_2$, if

$$\int H_1^2(v)dv \leq \int H_2^2(v)dv$$

and

$$\int v^2 H_1^2(v)dv \leq \int v^2 H_2^2(v)dv.$$

It is not difficult to show that the weight function $H$ minimizing

$$\int H^2(v)dv$$

subjects to the constraints

$$\int H(v)dv = 1, \ \int v^2 H(v)dv = c$$

is

$$H(v) = \begin{cases} -\dfrac{16}{9}a^3v^2 + a, & \text{for } |v| \leq \dfrac{3}{4a} \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

where $a = 3/4\sqrt{5c}$ (see e.g. Epanechnikov (1969) or Rosenblatt (1971)). On the other hand one can show, by the same method that subject to the restrains

$$\int H(v)dv = 1, \ \int H^2(v)dv = d,$$

the functional

$$\int v^2 H(v)dv$$

takes minimum for the weight function of the form (14) with $a = 5d/4$.

Therefore, since for every kernel $H$ there exists a kernel of the form (14) which is asymptotically not worse then $H$, then the weight function should be selected among a class of kernels defined by (14), $a \in (-\infty, \infty)$. Epanechnikov's (1969) discussion on the choice of the kernel is confined to the first of the two above minimization problems with $c = 1$.

## 5. Upper bounds for $\hat{S}_n(x)$ and $\tilde{S}_n(x)$

In this section we give upper bounds for $\hat{S}_n(x)$ and $\tilde{S}_n(x)$.

By the inequality

$$\left|\frac{L_n(x)}{M_n(x)} - R(x)\right| \leq \left|\frac{L_n(x)}{M_n(x)}\right| \left|\frac{M_n(x) - g(x)}{g(x)}\right| + \left|\frac{L_n(x) - R(x)g(x)}{g(x)}\right|,$$

the following two conditions

$$|L_n(x) - R(x)g(x)| \leq \frac{\varepsilon}{2+\varepsilon}|R(x)|g(x),$$

$$|M_n(x) - R(x)g(x)| \leq \frac{\varepsilon}{2+\varepsilon}g(x), \ \varepsilon > 0,$$

imply

$$\left|\frac{L_n(x)}{M_n(x)} - R(x)\right| \le \varepsilon|R(x)|.$$

By the above and Chebyshev's inequality

$$P\left\{\left|\frac{L_n(x)}{M_n(x)} - R(x)\right| > \varepsilon|R(x)|\right\}$$
$$\le \left(\frac{2+\varepsilon}{\varepsilon}\right)^2 \left[\frac{E\left[L_n(x) - g(x)R(x)\right]^2}{g^2(x)R^2(x)} + \frac{E\left[M_n(x) - g(x)\right]^2}{g^2(x)}\right]. \quad (15)$$

Let now

$$L_n(x) = \hat{R}(x)g(x),\ M_n(x) = \frac{1}{nh^p(n)}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h(n)}\right),$$

and $h(n) = kn^{-\alpha}$, $0 < \alpha < 1/p$. Since

$$E\left[M_n(x) - g(x)\right]^2 \sim d_1 k^4 n^{-4\alpha} + d_2 k^{-p} n^{-(1-\alpha p)},$$

where

$$d_1 = \frac{1}{4}\left(\sum_{i=1}^{p}\frac{\partial^2 g(x)}{\partial x_i^2}\right)^2 \int v^2 H(v)dv,$$
$$d_2 = \left[\int H^2(v)dv\right]^p g(x),$$

(see Epanechnikov (1969)), then, by (12) and (15), we get the following upper bound:

$$P\left\{\left|\hat{S}_n(x) - R(x)\right| > \varepsilon\,|R(x)|\right\} \le \left(\frac{2+\varepsilon}{\varepsilon}\right)^2 V_n, \quad (16)$$

where

$$V_n \sim \frac{k^4}{g^2(x)}\left(\frac{c_1}{R^2(x)} + d_1\right)n^{-4\alpha} + \frac{1}{k^p g^2(x)}\left(\frac{c_2}{R^2(x)} + d_2\right)n^{-(1-\alpha p)}.$$

Put, now

$$L_n(x) = \hat{R}(x)g(x),\ M_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^p(i)}K\left(\frac{x - X_i}{h(i)}\right).$$

By the method used in Lemmas 4 and 5 one can show, that for $h(n) = kn^{-\alpha}$, $0 < \alpha < \min(1/2, 1/p)$,

$$E\left[M_n(x) - g(x)\right]^2 \sim \frac{d_1 k^4}{(1-2\alpha)^2} n^{-4\alpha} + \frac{d_2}{k^p(1+\alpha p)} n^{-(1-\alpha p)}.$$

By this, (13) and (15), we get the upper bound for the recursive estimate:

$$P\left\{\left|\tilde{S}_n(x) - R(x)\right| > \varepsilon\,|R(x)|\right\} \le \left(\frac{2+\varepsilon}{\varepsilon}\right)^2 W_n, \quad (17)$$

where

$$W_n \sim \frac{k^4}{g^2(x)}\left(\frac{c_1}{R^2(x)} + \frac{d_1}{(1-2\alpha)^2}\right)n^{-4\alpha}$$
$$+ \frac{1}{k^p(1+\alpha p)g^2(x)}\left(\frac{c_2}{R^2(x)} + d_2\right)n^{-(1-\alpha p)}.$$

Both upper bounds (16) and (17) are minimized by $\alpha = \alpha^*$.

## References

[1] Ahmad, I.A. and P.A. Lin (1976). Nonparametric sequential estimation of a multiple regression function, *Bull. Math. Statist.* 17, 63-75.

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

[3] Epanechnikov, V.A. (1969). Nonparametric estimates of a multivariate probability density, *Theory Prob. Appl.* 14, 153-158.

[4] Nadaraya, E.A. (1964). On estimating regression, *Theory Prob. Appl.* 9, 141-142.

[5] Nadaraya, E.A. (1970). Remarks on non-parametric estimates for density function and regression curves, *Theory Prob. Appl.* 9, 134-137.

[6] Noda, K. (1976). Estimation of a regression function by the Parzen kernel-type density estimators, *Ann. Inst. Statist. Math.* 28, 221-234.

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

[8] Rosenblatt, M. (1971). Curve estimates, *Ann. Math. Statist.* 12, 1815-1842.

[9] Watson, G.S. (1964). Smooth regression, *Sankhyā Ser. A.* 26, 359-372.