# NONPARAMETRIC RECURSIVE VARIANCE ESTIMATION

by

**U. Stadtmüller** * **and A.B. Tsybakov** **

August 1992

**Abstract**

Consider an i.i.d sample for a random vector $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ . We are interested in recursive estimates of the regression function $f(x) = E(Y|X = x)$ and of the variance function $Var(Y|X = x)$ . We use recursive kernel estimates for both problems and prove a.s. convergence and mean square convergence with rates of convergence under appropriate conditions. The results are compared with nonrecursive estimators, which have recently been suggested by various authors.

AMS Classification: 62 G 05

Key words: nonparametric regression analysis, variance estimation, recursive estimation, kernel estimation, kernel estimates, speed of convergence.

* Universität Ulm, Abt. Mathematik III, Ulm, Germany

** CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium and Institute for Problems of Information Transmission, Russian Academy of Sciences, Russia.

## 1. Introduction

Let $(X,Y)$, $X \in I\!R^d$, $Y \in I\!R^1$ be random variables, with $E(Y^4) < \infty$, and let $f(x) = E(Y|X = x)$, and $\sigma^2(x) = \mathrm{Var}(Y|\ X = x)$ be the regression and conditional variance functions, $x \in I\!R^d$. Consider an i.i.d. sample $(X_i, Y_i)_{i=1}^n$ from $(X,Y)$, i.e.

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n\ , \tag{1.1}$$

with i.i.d. random variables $\varepsilon_i$ such that for all $x \in I\!R^d$

$$\begin{aligned} E(\varepsilon_1|X_1 = x) &= 0, \\ \mathrm{Var}(\varepsilon_1|X_1 = x) &= 1, \\ E(|\varepsilon_1|^4|X_1 = x) &= \mu_4(x) < \infty. \end{aligned} \tag{1.2}$$

Consider the problem of estimating $\sigma^2(x)$ from observations (1.1).

Assume that $\hat{f}_n(x)$ is the recursive estimator of $f(x)$ given by

$$\hat{f}_n(x) = \hat{f}_{n-1}(x) + \rho_n(Y_n - \hat{f}_{n-1}(x))H_n^{-d}K(\frac{X_n - x}{H_d}), \quad \hat{f}_0(\cdot) \equiv 0\ , \tag{1.3}$$

with some positive sequences $\{\rho_n\}, \{H_n\}$ and a kernel function $K : I\!R^d \to I\!R^1$. This estimator was proposed by Révész (1973). The convergence of $\hat{f}_n(x)$ in $L_2$ - norm and in $C$ - norm (with rates) was studied by Révész (1977); the a.s. pointwise convergence of a robust version of $\hat{f}_n(x)$ was considered by Tsybakov (1983). The estimator $\hat{f}_n(x)$ is not of main interest in this paper. We use it as a pilot one in the recursive variance estimation procedure which we define in the following way:

$$\hat{\sigma}_n^2(x) = \hat{\sigma}_{n-1}^2(x) + \gamma_n((Y_n - \hat{f}_{n-1}(x))^2 - \hat{\sigma}_{n-1}^2(x))h_n^{-d}k(\frac{X_n - x}{h_n})\ , \tag{1.4}$$

$$\hat{\sigma}_0^2(\cdot) \equiv 0\ , \qquad \text{where} \quad k : I\!R^d \to I\!R^1\ , \quad \text{and}$$

$$\gamma_n > 0, \quad \gamma_n \to 0, \quad \sum_{n=1}^{\infty} \gamma_n = \infty\ , \tag{1.5}$$

$$h_n > 0, \quad h_n \to 0, \quad nh_n^d \to \infty\ . \tag{1.6}$$

The smoothness assumptions are written in terms of Lipschitz continuity. We write $f \in \mathrm{Lip}_\alpha(I)$, $\alpha > 0$, $I$ some interval, if

$$(1.7) \qquad |f(x) - \sum_{|\nu| \le \lfloor \alpha \rfloor} f^{(\nu)}(x_0) \frac{(x - x_0)^\nu}{\nu!}| \le L|x - x_0|^{\alpha - \lfloor \alpha \rfloor} \quad \forall\, x,\ x_0 \in I,$$

where $\nu = (\nu_1, \ldots, \nu_d)$ is a multi-index, $\nu! = \nu_1! \ldots \nu_d!$, $|\nu| = \nu_1 + \ldots + \nu_d$, $f^\nu(x_0) = (\frac{\partial}{\partial x_1})^{\nu_1} \cdots (\frac{\partial}{\partial x_d})^{\nu_d} f(x)|_{x=x_0}$, $0 < L < \infty$ , and $\lfloor \alpha \rfloor$ is the maximal integer $< \alpha$ .

In the following we assume that $x \in I\!R^d$ is fixed, $\gamma, \varepsilon, L$ are fixed positive numbers, and that $\sigma^2, g \in \mathrm{Lip}_\gamma\ [x - \varepsilon, x + \varepsilon]$ , where $g(\cdot)$ is the marginal density of $X$ .

Assumptions on the kernel $k$ are the following: $k \in \mathcal{K}_\gamma$ , i.e. $k$ is bounded, and has support $[-1, 1]$ ,

$$(1.8) \qquad \int_{I\!R^d} k(v)dv = 1 \ ,$$

$$\int_{I\!R^d} k(v) v_1^{\nu_1} \ldots v_d^{\nu_d} dv = 0 \ ,$$

$$v = (v_1, \ldots, v_d), \quad \max(\lfloor \gamma \rfloor, 1) \le \nu_1 + \ldots + \nu_d \le \lfloor \gamma \rfloor \ .$$

Our main result gives the conditions of the mean squared and a.s. convergence of $\hat{\sigma}_n^2$ .

**Theorem 1.** Assume (1.1), (1.2), (1.5), (1.6), (1.8). Let $X$ have a density $g(\cdot)$ such that

$$(1.9) \qquad g(z) \ge g^* > 0, \quad z \in [x - \varepsilon, x + \varepsilon] \ ,$$

and $\sigma^2,\ g \in \mathrm{Lip}_\gamma[x - \varepsilon, x + \varepsilon]$ for some $\gamma,\ \varepsilon > 0$ .
Assume also that $\mu_4(z)$ is bounded uniformly in $z \in [x - \varepsilon, x + \varepsilon]$ .

(i) If

$$\begin{aligned} &\sum_{n=1}^{\infty} \gamma_n (f(x) - \hat{f}_{n-1}(x))^4 < \infty \qquad \text{(a.s.)} \ , \\ &\sum_{n=1}^{\infty} (\gamma_n h_n^{2\gamma} + \gamma_n h_n^{-d}) < \infty \ , \end{aligned} \tag{1.10}$$

then $\hat{\sigma}_n^2(x) \to \sigma^2(x)$ (a.s.), $n \to \infty$ .

(ii) If

$$E((f(x) - \hat{f}_{n-1}(x))^4) \to 0, \quad n \to \infty, \quad \gamma_n h_n^{-d} \to 0, \quad n \to \infty \ , \tag{1.11}$$

then $E((\hat{\sigma}_n^2(x) - \sigma^2(x))^2) \to 0, \quad n \to \infty$ .

(iii) If

$$\begin{aligned} &E((f(x) - \hat{f}_n(x))^4) = \mathcal{O}(n^{-4\delta}), \quad \delta > 0, \quad \gamma_n = \frac{c^*}{n} \ , \\ &h_n = h^* n^{-\frac{1}{2\gamma+d}}, \ h^* > 0, \ c^* > (2g^*)^{-1} \min(2\gamma/(2\gamma+d), 4\delta), \end{aligned} \tag{1.12}$$

then

$$E((\hat{\sigma}_n^2(x) - \sigma^2(x))^2) = \mathcal{O}(\max\{n^{-2\gamma/(2\gamma+d)}, n^{-4\delta}\}), \quad n \to \infty \ . \tag{1.13}$$

The proof of Theorem 1 is given in Section 2.

**Remarks:** For parametric approaches to estimate the variance function we refer to the book of Carroll and Ruppert (1988). Let us compare our results with other nonparametric approaches to estimate the variance function. In Hall and Carroll (1989) the following nonrecursive approach is suggested. They consider the case $d = 1$ and estimate the regression function $f$ by a linear estimator, e.g. a kernel estimator. Choosing all parameters in an optimal way, they obtain for $f \in \text{Lip}_\beta[x - \varepsilon, x + \varepsilon]$ :

$$E((\hat{f}_n(x) - f(x))^2) = \mathcal{O}(n^{-2\beta/(2\beta+1)}) \ .$$

Then they consider empirical residuals $\hat{\varepsilon}_i = Y_i - \hat{f}(X_i), \quad 1 \le i \le n$ , and estimate $\sigma^2(x)$ by a kernel estimator again from the data $\hat{\varepsilon}_i^2$ resulting in an estimator $\hat{\sigma}^2(x)$ . They show that by this procedure one can obtain the rate

$$E\left((\hat{\sigma}^2(x) - \sigma^2(x))^2\right) = \mathcal{O}(\max(n^{-4\beta/(2\beta+1)}, n^{-2\gamma/(2\gamma+1)})) \tag{1.14}$$

and that this rate is optimal. Comparing (1.14) with (1.13) it can be seen that provided that we can achieve $\delta = \beta/(2\beta+1)$ (see Theorem 2 below) our recursive estimator has the optimal rate . But naturally a recursive approach is simpler to handle if one has increasing sample size. Furthermore our Theorem 1 treats the multivariate case as well.

In Müller and Stadtmüller (1991) a more general approach has been considered, for the case of fixed design regression models. There $\sigma^2(x)$ is estimated by a quadratic form in the variables $Y_i$ . It is shown again that essentially the same rate as above can be obtained, but in some cases it can be even improved somewhat by that general approach. Observe that the estimator of Hall and Carroll as well as our recursive procedure are all quadratic in the data $\{Y_i\}$ .

Härdle and Tsybakov (1988) introduced a class of nonparametric variance estimates, non-quadratic in $\{Y_j\}$ , related to M-estimates of scale. They showed pointwise consistency and asymptotic normality of these estimators for the case of $\gamma = \beta = 1,\ d \geq 1$ , under general conditions on the distribution of $\varepsilon_j$ .

For applications of variance estimation see e.g. Robinson (1987) or Müller and Stadtmüller (1987).

For our estimator the following question remains. Can we achieve $\delta = \beta/(2\beta+1)$ if $f \in \text{Lip}_\beta[x-\varepsilon, x+\varepsilon]$ , i.e. can the regression function be estimated sufficiently good? Under our conditions and for the algorithm (1.3) the relations

$$E((f(x) - \hat{f}_{n-1}(x))^4) \to 0, \quad n \to \infty \,, \tag{1.15}$$

and

$$E((f(x) - \hat{f}_{n-1}(x))^4) = \mathcal{O}(n^{-4\delta}), \quad \delta > 0 \tag{1.16}$$

are not given with literature. This gap will be closed by the next result.

**Theorem 2.** Assume (1.1) and (1.2). Let $K \in \mathcal{K}_\beta$ and let $X$ have a density $g(\cdot)$ such that

$$g(z) \geq g^* > 0, \quad z \in [x-\varepsilon, x+\varepsilon] \,,$$

and $f, g \in \text{Lip}_\beta[x-\varepsilon, x+\varepsilon]$ , for some $\beta, \varepsilon > 0$ .

Assume also that $\sigma^2(z)$ and

$$E(|\varepsilon_1|^r \,|X_1 = z) \,.$$

are bounded uniformly in $z \in [x-\varepsilon, x+\varepsilon]$ for some $r \geq 2$ .

(i) If

$$\sum_{n=1}^{\infty} \rho_n [H_n^{\beta r} + (\rho_n H_n^{-d})^{r/2}] < \infty \tag{1.17}$$

then

$$\sum_{n=1}^{\infty} \rho_n |f(x) - \hat{f}_{n-1}(x)|^r < \infty \quad \text{(a.s.)} \,,$$

and if, moreover,

$$\sum_{n=1}^{\infty} \rho_n = \infty \tag{1.18}$$

then $\hat{f}_{n-1}(x) \xrightarrow{a.s.} f(x), \;\; n \to \infty$ .

(ii) If $H_n \to 0, \;\; \rho_n H_n^{-d} \to 0$ and (1.18) is satisfied then

$$E(|\hat{f}_{n-1}(x) - f(x)|^r) \to 0, \quad n \to \infty \,.$$

(iii) If $\rho_n = \rho^* n^{-1}, \;\; H_n = H^* n^{-\frac{1}{2\beta+d}}, \;\; H^* > 0, \;\; \rho^* > \beta/(g^*(2\beta+d))$ then

$$E(|\hat{f}_{n-1}(x) - f(x)|^r) = \mathcal{O}(n^{-\frac{r\beta}{2\beta+d}}), \quad n \to \infty \,.$$

**Remark:** Theorems 1 and 2 imply in particular that under the assumptions as in Hall and Carroll (1989) we can achieve the optimal rate (1.14) by our recursive estimator.

## 2. Proofs.

**Proof of Theorem 1.** In the proofs assume that $x \in I\!R^d$ is fixed and use the following notation

$$s_n = \hat{\sigma}_n^2(x), \quad f_n = \hat{f}_n(x), \quad s = \sigma^2(x), \quad f = f(x)\ ,$$
$$U_n = s_n - s, \quad \mathcal{Z}_n = (X_i, Y_i)_{i=1}^n, \quad \xi_n = Y_n - E(Y_n|X_n)\ .$$

Later we assume that $n$ is large enough, so that $h_n < \varepsilon$ . It follows from (1.4) that

$$E(U_n^2|\mathcal{Z}_{n-1}) = U_{n-1}^2 + 2U_{n-1}\gamma_n h_n^{-d}[E((Y_n - f_{n-1})^2 k(\frac{X_n - x}{h_n})|\mathcal{Z}_{n-1}) - \tag{2.1}$$
$$-s_{n-1}E(k(\frac{X_n - x}{h_n}))] +$$
$$+\gamma_n^2 h_n^{-2d} E([(Y_n - f_{n-1})^2 - s_{n-1}]^2 k^2(\frac{X_n - x}{h_n})|\mathcal{Z}_{n-1})\ .$$

Now, since $E(\xi_n|X_n) = 0$ and $E(\xi_n^2|X_n) = \sigma^2(X_n)$ we have

$$h_n^{-d}[E((Y_n - f_{n-1})^2 k^2(\frac{X_n - x}{h_n})|\mathcal{Z}_{n-1}) - s_{n-1}E(k(\frac{X_n - x}{h_n}))] = \tag{2.2}$$
$$= h_n^{-d}[E[((f - f_{n-1})^2 - 2\xi_n(f - f_{n-1}) + \xi_n^2)k(\frac{X_n - x}{h_n})|\mathcal{Z}_{n-1}] -$$
$$-s_{n-1}E(k(\frac{X_n - x}{h_n}))] =$$
$$= (f - f_{n-1})^2 h_n^{-d} E(k(\frac{X_n - x}{h_n})) +$$
$$+h_n^{-d}\int (\sigma^2(z) - s_{n-1})k(\frac{z - x}{h_n})g(z)dz\ \ .$$

Since the kernel $k$ satisfies (1.8) and $g \in \mathrm{Lip}_\gamma[x - \varepsilon, x + \varepsilon]$ we have

$$h_n^{-d}E(k(\frac{X_n - x}{h_n})) = \int k(u)g(x + uh_n)du = g(x) + \mathcal{O}(h_n^\gamma), \quad n \to \infty\ , \tag{2.3}$$

and similarly, since $(\sigma^2(\cdot)-\sigma^2(x))g(\cdot)\in \mathrm{Lip}_\gamma[x-\varepsilon,x+\varepsilon]$ ,

$$h_n^{-d}\int(\sigma^2(z) - \sigma^2(x))k(\frac{z-x}{h_n})g(z)dz = ((\sigma^2(z)-\sigma^2(x))g(z))|_{z=x} + \tag{2.4}$$

$$+ \quad \mathcal{O}(h_n^\gamma)=\mathcal{O}(h_n^\gamma),\ n\to\infty\,.$$

By substitution of (2.3), (2.4) into (2.2) we find

$$h_n^{-d}[E((Y_n-f_{n-1})^2k^2(\frac{X_n-x}{h_n})|\mathcal{Z}_{n-1}) - s_{n-1}E(k(\frac{X_n-x}{h_n}))] = \tag{2.5}$$

$$=[(f-f_{n-1})^2+(s-s_{n-1})]g(x)+\mathcal{O}(h_n^\gamma)(1+(f-f_{n-1})^2+(s-s_{n-1}))=$$

$$=-U_{n-1}(g(x)+\mathcal{O}(h_n^\gamma))+c_1(f-f_{n-1})^2+\mathcal{O}(h_n^\gamma),\ \ n\to\infty\,.$$

Here and later $c_i,\ \ i=1,2,\dots$ are positive constants. Next,

$$h_n^{-d}E([(Y_n-f_{n-1})^2-s_{n-1}]^2k^2(\frac{X_n-x}{h_n})|\mathcal{Z}_{n-1}) = \tag{2.6}$$

$$=h_n^{-d}E([(f-f_{n-1})^2-2\xi_n(f-f_{n-1})+(\xi_n^2-s_{n-1})]^2\ \ k^2(\frac{X_n-x}{h_n})|\mathcal{Z}_{n-1})\ \le$$

$$\le h_n^{-d}[(f-f_{n-1})^4E(k^2(\frac{X_n-x}{h_n}))+4(f-f_{n-1})^2E(\xi_n^2k^2(\frac{X_n-x}{h_n}))+$$

$$+E((\xi_n^2-s_{n-1})^2k^2(\frac{X_n-x}{h_n})|\mathcal{Z}_{n-1})].$$

Here

$$h_n^{-d}E(k^2(\frac{X_n-x}{h_n}))=\int k^2(u)g(x+uh_n)du=\mathcal{O}(1),\quad n\to\infty\,, \tag{2.7}$$

and similarly

$$h_n^{-d}E(\xi_n^2k^2(\frac{X_n-x}{h_n}))=\mathcal{O}(1),\quad n\to\infty. \tag{2.8}$$

Also, since $\sigma^2(X_n) = E(\xi_n^2|X_n), \quad \xi_n = Y_n - E(Y_n|X_n)$

$$
\begin{aligned}
(2.9) \qquad h_n^{-d} \;& E((\xi_n^2 - s_{n-1})^2 k^2(\tfrac{X_n - x}{h_n})|\mathcal{Z}_{n-1}) \leq \\
\leq \;& 2h_n^{-d}[E((\xi_n^2 - \sigma^2(X_n))^2 k^2(\tfrac{X_n - x}{h_n})|\mathcal{Z}_{n-1}) + \\
+ \;& E((\sigma^2(X_n) - s_{n-1})^2 k^2(\tfrac{X_n - x}{h_n})|\mathcal{Z}_{n-1})] \leq \\
\leq \;& 2[h_n^{-d} \textstyle\int \mu_4(z) k^2(\frac{z-x}{h_n}) g(z) dz + \\
+ \;& 2(s - s_{n-1})^2 h_n^{-d} \textstyle\int k^2(\frac{z-x}{h_n}) g(z) dz + \\
+ \;& 2h_n^{-d} \textstyle\int (\sigma^2(z) - \sigma^2(x))^2 k^2(\frac{z-x}{h_n}) g(z) dz] \leq \\
\leq \;& c_2[1 + (s - s_{n-1})^2] = c_2[1 + U_{n-1}^2]
\end{aligned}
$$

where we used the fact that $\sigma^2(x)$ is uniformly continuous in $[x - \varepsilon, x + \varepsilon]$, and thus the last integral is $o(1)$ as $n \to \infty$.

Substituting now (2.7), (2.8), (2.9) into (2.6) and using the inequality $2(f_{n-1} - f)^2 \leq 1 + (f_{n-1} - f)^4$ we find:

$$
\begin{aligned}
(2.10) \qquad & h_n^{-d} E([(Y_n - f_{n-1})^2 - s_{n-1}]^2 k^2(\frac{X_n - x}{h_n})|\mathcal{Z}_{n-1}) \leq \\
& \leq c_3[(U_{n-1}^2 + 1) + (f_{n-1} - f)^4] \,.
\end{aligned}
$$

Now we use (2.5) and (2.10) in order to get the following estimation for (2.1):

$$
\begin{aligned}
E(U_n^2|\mathcal{Z}_{n-1}) \;& = \; U_{n-1}^2 + 2U_{n-1}\gamma_n[-U_{n-1}(g(x) + \mathcal{O}(h_n^\gamma)) + c_1(f - f_{n-1})^2 + \mathcal{O}(h_n^\gamma)] \\
& + \; c_3\gamma_n^2 h_n^{-d}(1 + U_{n-1}^2 + (f_{n-1} - f)^4) \\
& \leq \; U_{n-1}^2(1 - 2\gamma_n(g(x) + \mathcal{O}(h_n^\gamma)) + c_3\gamma_n^2 h_n^{-d}) \\
& + \; c_4 U_{n-1}\gamma_n((f - f_{n-1})^2 + h_n^\gamma) \\
& + \; c_3\gamma_n^2 h_n^{-d}(1 + (f_{n-1} - f)^4) \,.
\end{aligned}
$$

But observe that $U_{n-1} \leq \frac{1}{2}(q + \frac{1}{q}U_{n-1}^2)$, for all $q > 0$. Now we choose $q = ((f - f_{n-1})^2 + h_n^\gamma)(2g(x)q_0)^{-1}c_4$ for some $q_0 \in (0, 1)$. Then

$$(2.11)\quad E(U_n^2|\mathcal{Z}_{n-1}) \le U_{n-1}^2[1-2\gamma_n(g(x)+\mathcal{O}(h_n^\gamma))+g(x)q_0\gamma_n+c_3\gamma_n^2h_n^{-d}]+$$
$$+c_6\gamma_n((f-f_{n-1})^2+h_n^\gamma)^2\frac{1}{2g(x)q_0}+c_3\gamma_n^2h_n^{-d}(1+(f_{n-1}-f)^4),$$

and we obtain

$$(2.12)\qquad E(U_n^2|\mathcal{Z}_{n-1}) \le$$
$$\le U_{n-1}^2[1-2\gamma_n g(x)(1-q_0/2)]+c_7\gamma_n[(f-f_{n-1})^4+h_n^{2\gamma}+\gamma_n h_n^{-d}]$$

where we used the fact that $h_n^\gamma \to 0,\ \gamma_n h_n^{-d} \to 0, \text{as}\ n\to\infty$ .

**For part (i)**

Using (2.12) and a result of Robbins and Siegmund (consult e.g. Duflo (1990), Theorem 1.III.12) we find that the conditions (1.5) and (1.10) are sufficient for the convergence $U_n^2 \xrightarrow{\text{a.s.}} 0,\ n\to\infty$ .

**For parts (ii) and (iii)**

Take expectations on both sides of (2.12) and denote $E(U_n^2)=v_n$ . We have

$$(2.13)\ v_n \le v_{n-1}(1-2\gamma_n g(x)(1-q_0/2))+c_7\gamma_n[E((f-f_{n-1})^4)+h_n^{2\gamma}+\gamma_n h_n^{-d}].$$

Denoting $a_n^4=(f_{n-1}-f)^4$ we can use a result stated in Polyak (1987), Ch.2, Sec. 2 under the conditions

$$\sum_n \gamma_n=\infty,\quad a_n,\gamma_n,h_n\to 0,\quad \gamma_n h_n^{-d}\to 0$$

to obtain $v_n\to 0,\ n\to\infty$ i.e. the mean-square convergence of our estimator. If $\gamma_n=\frac{c^*}{n}$ then it follows from (2.13) that

$$v_n \le v_{n-1}(1-2c^*g^*(1-q_0/2)n^{-1})+c_8n^{-1}(a_n^4+h_n^{2\gamma}+(nh_n^d)^{-1})\ .$$

By Chung's Lemma (see e.g. Polyak (1987), Ch.2, Sec. 2 or Wasan (1969))

using $a_n = c_9 n^{-\delta},\ \ \delta > 0,\ \ h_n = c_{10} n^{-1/(2\gamma+d)},\ \ c^* > (2g^*)^{-1} \min(\frac{2\gamma}{2\gamma+d}, 4\delta)$ we get

$$v_n = \mathcal{O}(n^{-\min(\frac{2\gamma}{2\gamma+d}, 4\delta)}), \quad n \to \infty \ .$$

**Proof of Theorem 2.** For any $r \geq 2, z, \Delta \in I\!R^d$ we have using Taylor's formula ( $(z, \Delta)$ denotes the canonical inner product)

$$|z+\Delta|^r \leq |z|^r + r|z|^{r-2}(z,\Delta) + \frac{1}{2} r(r-1)(|z|+|\Delta|)^{r-2}|\Delta|^2. \tag{2.14}$$

Denote $V_n = |f_n - f|^r$ . By (1.3), (2.14) have

$$V_n \leq V_{n-1} + r V_{n-1}^{(r-2)/r}((f_{n-1}-f), \Delta_n) + c_{12}[V_{n-1}^{(r-2)/r} + |\Delta_n|^{r-2}]|\Delta_n|^2 \tag{2.15}$$

where

$$\Delta_n = \rho_n (Y_n - f_{n-1}) H_n^{-d} K(\frac{X_n - x}{H_n}) \ .$$

Now

$$E(\Delta_n | \mathcal{Z}_{n-1}) = \rho_n H_n^{-d}[(f(X_n) - f(x)) + (f(x) - f_{n-1}(x))] E(K(\frac{X_n - x}{H_n})) \ .$$

Since $K \in \mathcal{K}_\beta$ and $f, g \in \text{Lip}_\beta[x-\varepsilon, x+\varepsilon]$ we obtain (cf. (2.3), (2.4))

$$E(\Delta_n | \mathcal{Z}_{n-1}) = \rho_n (f - f_{n-1})(g(x) + \mathcal{O}(H_n^\beta)), \quad n \to \infty \ . \tag{2.16}$$

Next,

$$\begin{aligned} E!(|\Delta_n|^t | \mathcal{Z}_{n-1}) \leq \qquad & (2.17) \\ \leq c_{13} \rho_n^t H_n^{-dt} E([|f - f_{n-1}|^t + E(|\xi_n|^t | \mathcal{Z}_{n-1})] | K(\frac{X_n - x}{H_n})|^t) \leq \\ \leq c_{14} \rho_n^t H_n^{d(1-t)} [V_{n-1}^{t/r} + 1], \quad \text{for } 1 \leq t \leq r \ , \end{aligned}$$

where $c_{13}, c_{14}$ depend only on $t$ . Here we used the a.s. boundedness of $E(|\xi_n|^t | \mathcal{Z}_{n-1}), t \leq r$ .

Observe that $E((f_{n-1}-f,\Delta_n)|\mathcal{Z}_{n-1}) \le |f_{n-1}-f|\cdot|E(\Delta_n|\mathcal{Z}_{n-1})|$ .
By substitution of (2.16), (2.17) into (2.15) we find

$$\begin{aligned} E(V_n|\mathcal{Z}_{n-1}) &\le V_{n-1} - r\rho_n V_{n-1} g(x) + c_{15} r \rho_n V_{n-1}^{(r-1)/r} H_n^{\beta} + \\ &+ c_{16}[V_{n-1}^{(r-2)/r}\rho_n^2 H_n^{-d}(V_{n-1}^{2/r}+1) + \\ &+ \rho_n^r H_n^{d(1-r)}(V_{n-1}+1)] \le \\ &\le V_{n-1}[1-\rho_n(rg^* - c_{16}\rho_n H_n^{-d} - c_{16}(\rho_n H_n^{-d})^{r-1})] + \\ &+ c_{17}\rho_n[H_n^{\beta} V_{n-1}^{(r-1)/r} + \rho_n H_n^{-d} V_n^{(r-2)/r} + (\rho_n H_n^{-d})^{r-1}] \, . \end{aligned}$$

Using the elementary inequality

$$V_n^{\tau} \le (1-\tau)a^{\tau} + \tau a^{\tau-1} V_n, \quad \forall\, a > 0 \, ,$$

for $\tau = (r-1)/r$ and for $\tau = (r-2)/r \ge 0$ and choosing $a$ so that $c_{17} H_n^{\beta}\tau a^{\tau-1} = \tau_0 > 0$ in the first case, and $c_{17}\rho_n H_n^{-d}\tau a^{\tau-1} = \tau_0 > 0$ in the second case, where $\tau_0$ is a constant to be fixed later, we find

$$E(V_n|\mathcal{Z}_{n-1}) \le \tag{2.18}$$

$$\le V_{n-1}[1-\rho_n(rg^* - 3\tau_0)] + c_{18}\rho_n[H_n^{\beta r} + (\rho_n H_n^{-d})^{r/2} + (\rho_n H_n^{-d})^{r-1}]$$

if $n$ is large enough, so that

$$c_{16}\rho_n H_n^{-d} + c_{16}(\rho_n H_n^{-d})^{r-1} < \tau_0 \, .$$

Choosing $\tau_0$ in (2.18) small enough and using the theorem of Robbins–Siegmund (Duflo (1990), Theorem 1.III.12) we get that the conditions

$$\sum_{n=1}^{\infty} \rho_n [H_n^{\beta r} + (\rho_n H_n^{-d})^{r/2}] < \infty \, , \tag{2.19}$$

$$\sum_{n=1}^{\infty} \rho_n = \infty \tag{2.20}$$

are sufficient for the convergence $V_n \to 0, \quad n \to \infty$ , (a.s.). Also we can rewrite (2.18) as

$$E(V_n|\mathcal{Z}_{n-1}) \leq V_{n-1} + \chi_n - \eta_n$$

where

$$\begin{aligned} \chi_n &= c_{18}\rho_n[H_n^{\beta r} + (\rho_n H_n^{-d})^{r/2} + (\rho_n H_n^{-d})^{r-1}] \,, \\ \eta_n &= V_n \rho_n (r g^* - 3\tau_0) \,, \end{aligned}$$

and the theorem of Robbins–Siegmund entails that under (2.19)

$$\sum_{n=1}^{\infty} V_n \rho_n = \sum_{n=1}^{n} \eta_n < \infty \qquad \text{(a.s.)}$$

if $rg^* > 3\tau_0$ . If $\rho_n = \gamma_n, \quad r = 4$ , then we get the first inequality in (1.10). This proves the part (i) of Theorem 2.

Now, take expectations of both sides of (2.18). By Polyak (1987), Ch. 2, Sec. 2, we have that $E(V_n) \to 0, \quad n \to \infty$ , if $H_n \to 0, \quad \rho_n H_n^{-d} \to 0$ , and if (2.20) is satisfied. This gives the part (ii) of Theorem 2. If $\rho_n = \rho^* n^{-1}$ , $H_n = c_{19} n^{-\frac{1}{2\beta+d}}$ ,

$$\rho^* > \frac{\beta}{g^*(2\beta + d)}$$

then, using Polyak (1987), Ch.2, Sec. 2, we obtain

$$E(V_n) = \mathcal{O}(n^{-\frac{r\beta}{2\beta+d}}), \quad n \to \infty.$$

This competes the proof of Theorem 2.

## References

**Carroll R.J., Ruppert D.** (1988), Transformation and Weighting in Regression. Chapman and Hall, London.

**Duflo M.** (1990), Methodes recursives aléatoires. Masson, Paris.

**Hall P., Carroll R.J.** (1989), Variance function estimation in regression: the effect of estimating the mean. J. R. Statist. Soc. B 51, 3–14.

**Härdle W., Tsybakov A.B.** (1988), Robust nonparametric regression with simultaneous scale curve estimation. Ann. Statist. 16, 120–135.

**Müller H.-G., Stadtmüller U.** (1991), On variance function estimation with quadratic forms; To appear in: J. Stat. Plann Inf.

**Müller H.-G., Stadtmüller U.** (1987), Estimation of heteroscedasticity in regression analysis, Ann. Statist. 15, 610–625.

**Polyak B.T.** (1987), Introduction to Optimization. Optim. Software, N.Y.

**Révész P.** (1973), Robbins–Monro procedure in a Hilbert space and its application in the theory of learning processes. I. - Stud. Sci. Math. Hungar., 8, 391–398.

**Révész P.** (1977), How to apply the method of stochastic approximation in the nonparametric estimation of regression function. Math. Operationsforsh. Statist., Ser. Statist., 8, 119–126.

**Robinson P.M.** (1987), Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. Econometrika 55, 875–891.

**Tsybakov A.B.** (1983), Convergence of nonparametric robust algorithms of reconstruction of functions. Automation and Remote Control, 44, n.12, part 1, 1582–1591.

**Wasan M.T.** (1969), Stochastic Approximation. Cambridge University Press, Cambridge.