Revista de la
Unión Matemática Argentina
Vol. 35 1990

ASYMPTOTIC DISTRIBUTION AND STRONG ORDER OF CONVERGENCE OF ROBUST NONPARAMETRIC REGRESSION ESTIMATES

GRACIELA BOENTE - RICARDO FRAIMAN

Summary

In this paper, strong order of convergence and the finite dimensional asymptotic distribution of nearest neighbor and of nearest neighbor with kernel regression estimates is obtained when the response variables are bounded. These results are applied for deriving the asymptotic normality and orders of convergence for robust nonparametric regression estimates.

1. Introduction

Let (X,Y), $(X_1,Y_1),\ldots,(X_n,Y_n)$ be independent, identically distributed random vectors, $Y \in R, X \in R^p$. As is well known, nonparametric estimators of the regression function $r(x) = E(Y|X = x)$ can be written as $r_n(x) = \sum_{i=1}^n W_{ni}(x)Y_i$, where $W_{ni}(x) = W_{ni}(x, X_1,\ldots,X_n)$ is a probability weight function, i.e., $W_{ni}(x) \geq 0$ and $\sum_{i=1}^n W_{ni}(x) = 1$.

Nearest neighbor methods were introduced by Cover (1968) and the weights are defined as follows. Rank the (X_i,Y_i), $1 \leq i \leq n$, according to increasing values of $\|X_i - x\|$ and obtain a vector of indices $(R_1,\ldots,R_n)$ where X_{R_i} is the ith nearest neighbor of x among $X_1,\ldots,X_n$. Let $\{v_{ni}\, , \, i \geq 1\}$ be a sequence of real numbers such that $v_{n1} \geq v_{n2} \geq \ldots \geq v_{nn} \geq 0$, $v_{ni} = 0 \quad i > n$, and $\sum_{i=1}^n v_{ni} = 1$, then $W_{nR_i}(x) = v_{ni}$. The consistency properties of $r_n(x)$ for different choices of the vector $(v_{n1},\ldots,v_{nn})$ were studied by Cover (1968), Stone (1977), Devroye (1978) and (1981). Devroye (1982) gives necessary and sufficient conditions for pointwise convergence.

Nearest neighbor with kernel methods were introduced for density estimation by Rosenblatt (1979) and by Mack and Rosenblatt (1979) and adapted to regression models by Collomb (1980). The weight function can be written as $W_{ni}(x) = K((X_i - x)/H_n)/\sum_{j=1}^n K((X_j - x)/H_n)$ where $H_n = H_n(x) = \|X_{R_k} - x\|$, $k = k_n$ is a sequence of integers and $K : R^p \to R$ is a nonnegative integrable function. Mack (1981) derived the asymptotic normality of nearest neighbor with kernel estimates for kernels supported on the unit ball; in particular, the asymptotic distribution of uniform k-nearest neighbor estimates, defined by $v_{ni} = 1/k$ for $1 \leq i \leq k \quad v_{ni} = 0 \quad i > k$, is derived.

The first reference for kernel estimates are Nadaraya (1964) and Watson's (1964) papers. Further developping was done, for instance, by Devroye and Wagner (1980 a) and b)), Spiegelman and Sacks (1980), Grebliki, Krzyźak and Pawlak (1984) and Gyorfi (1981).In this case, $W_{ni}(x) = K((X_i - x)/h)/\sum_{j=1}^{n} K((X_j - x)/h)$ with $h = h_n$ a sequence of nonnegative real numbers.

In Boente and Fraiman (1989) a robust location functional was introduced and strongly consistent, robust nonparametric estimates of the regression function were obtained by applying nonparametric distribution estimates to the functional equation.

Let Ψ be a bounded, strictly increasing real function, denote by $F(y|X = x)$ a regular version of the conditional distribution function of $Y|X = x$ and define $g(x)$ as the solution of

$$\int \Psi((y - g(x))/s(x))\, dF(y|X = x) = 0 \tag{1.1}$$

where $s(x)$ is any robust scale measure, for instance, $s(x) = MAD_c(x) = med(|Y - m(x)|\,|X = x)$ where $m(x) = med(Y|X = x)$ is the median of the conditional distributional function. If F is symmetric around $r(x)$ and Ψ is odd $g(x) = r(x)$.

The robust nonparametric estimate is defined as the unique solution $g_n(x)$ of

$$\sum_{i=1}^{n} W_{ni}(x)\, \Psi((Y_i - g_n(x))/s_n(x)) = 0 \tag{1.2}$$

where the scale measure is, for instance $s_n(x) = med(|Y - m_n(x)|\,|X = x)$ and the medians are evaluated corresponding to the conditional empirical distribution function;

$$F_n(y|X = x) = \sum_{i=1}^{n} W_{ni}(x)\, I_A(Y_i) \tag{1.3}$$

where $A = (-\infty, y]$ and I_A denotes the indicator function of the set A . More generally, any robust estimator consistent to $s(x)$ can be used.

In Section 2 strong order of consistency are obtained. In Section 3 the asymptotic normality of both, linear and robust, estimates is stated. The bias of the robust estimate is the same as for its linear relative and the relationship between the asymptotic variances is the same as for the usual robust location estimates. In Section 4, proofs are given.

2. Strong convergence rates

In Boente and Fraiman (1989) the unicity of solution of the equation (1.1), the weak continuity of the functional defined through (1.1) and strong consistency of $g_n(x)$ was obtained under:

H1. $\Psi : R \to R$ is a strictly increasing, bounded and continuous function such that $lim_{t\to+\infty}\Psi(t) = a > 0$ and $lim_{t\to-\infty}\Psi(t) = b < 0$.

H2. There exists a sequence $\{c_n : n \geq 1\}$ of real numbers such that $c_n \geq 0$, $c_n log\ n \to 0$, $n\, c_n \to \infty$ as $n \to \infty$, for which $max_{1\leq j\leq n} W_{nj}(x, X_1, \ldots, X_n) \leq c_n$ a.s. for almost all $x(P_X)$.

H3. There exists a random variable $\mathcal{K}_n$ and a real number $c > 0$ verifying $\sum_{i\in I_{nR_{\mathcal{K}_n}}} W_{ni}(x, X_1, \ldots, X_n) \to 0$ as $n \to \infty$ a.s. $sup_n(c_n\mathcal{K}_n) \leq c$ a.s. for almost all $x(P_X)$.

In order to obtain strong convergence rates for the robust estimators of the regression function we will need sôme additional regularity conditions.

H4. The vector X has a density f continuous and positive at x.

H5. $F(y|X=x)$ is symmetric around $g(x)$.

H6. $F(y|X=x)$ is continuous as a function of y and Lipschitz in x uniformly in y, i.e., there exists $\delta > 0$ and $c > 0$ such that $\|u-x\| < \delta \Rightarrow |F(y|X=x) - F(y|X=u)| \leq c\|u-x\|$ for all y.

H7. There exists $c > 0$ such that $P(\theta_n^{-1} \sum_{i \in I_{nR_{K_n}}} W_{ni}(x) \leq c) = 1$ where $\theta_n = (c_n log\, n)^{1/2}$.

H8. There exists $a_0 > 0$ such that $a_0 < c_n^{1+\frac{2}{d}} n^{\frac{2}{d}} log\, n$ for all n.

Remark 2.1. As noted in Boente and Fraiman (1989) kernel weights do not verify H2. However, the strong consistency of $g_n(x)$ can be obtained from Theorem 2 of Greblicki, Krzyżak and Pawlak [13] if K and h_n satisfy the following assumptions:

i) $h_n \to 0$ and $nh_n^d / log\, n \to \infty$ as $n \to \infty$.

ii) There exist positive constants, $r, c_1, c_2, c_3,$ and a bounded Borel function H decreasing on $(0, +\infty)$ such that $c_1 H(\|x\|) \leq K(x) \leq c_2 H(\|x\|), c_3 I_{\|x\| \leq r}(x) \leq K(x)$ and $t^d H(t) \to 0$ as $t \to \infty$.

Assumption H7 is fulfilled for any $k-NN$ weight function, in particular for the k-nearest neighbor with kernel if there exist positive constants c_1 and c_2 such that

$$c_1 I_{\|t\| \leq 1}(u) \leq K(u) \leq c_2 I_{\|t\| \leq 1}(u) \tag{2.1}$$

A nearest neighbor satisfies H7 if $\theta_n^{-1} \sum_{i > k_n} v_{ni}$ is bounded.

Lemma 2.1. *Under H2 to H4 and H6 to H8 we have that*

$$\theta_n^{-1} \sup_y |F_n(y|X=x) - F(y|X=x)| = 0(1) \quad a.s.$$

The proof may be found in section 4.

Theorem 2.3 of Boente and Fraiman (1989) together with Lemma 2.1 implies the following result.

Theorem 2.1. *If ψ is odd, continuously differentiable with derivative ψ' positive and bounded, H1 to H8 imply that*

$$\theta_n^{-1}(g_n(x) - g(x)) = 0(1) \quad a.s.$$

Remark 2.2. The conclusion of Theorem 2.1 also holds for kernel weights under H4, H5 and H6 provided that the sequence $\{h_n : n \geq 1\}$ and the kernel K satisfy the conditions given in Remark 2.1 and the following additional conditions: $h_n \theta_n^{-1} \leq A \leq \infty$ for all n where $\theta_n = (\log n / nh_n^d)^{1/2}$ and $t^{d+2} H(t)$ is bounded.

3. Asymptotic Distribution

Finally, we will show how to derive the asymptotic distribution of the robust regression estimates by reducing the problem to obtain the asymptotic distribution of the "classical" nonparametric regression estimates for bounded variables. We will need the following additional assumption on the function ψ.

H9. The function ψ is twice continuously differentiable with second derivative ψ'' verifying that there exist positive constants c, M and ϵ such that $|\psi''(t)| \le c\,|t|^{-(2+\epsilon)}$ for $|t| \ge M$.

H10. $E\{\psi'[(Y - g(x))/s(x)]|X = x\} \neq 0$.

We will denote by $\psi_\sigma(t) = \sigma\psi(t/\sigma)$.

Lemma 3.1. *Under H1, H5 and H9, if there exists a real constant* $c > 0$ *and a sequence of positive numbers* $\{c_n : n \ge 1\}$ *such that* $c_n^{-1}E[\sum_{i=1}^n W_{ni}^2(x, X_1, \ldots, X_n)] \le c$, *we have that*

$$c_n^{1/2} \sum_{i=1}^{n} W_{ni}(x)[\psi_{\sigma_n}(Y_i - g(x)) - \psi_\sigma(Y_i - g(x))] \to 0 \tag{3.1}$$

in probability for any sequence $\sigma_n = \sigma_n(x)$ *such that* $\sigma_n(x) \to \sigma(x) = \sigma > 0$ *in probability.*

Proof. For each fixed x we define

$$H_t(u) = \psi_{\sigma+t}(u) - \psi_\sigma(u),\ I_t(u) = \psi_{\sigma-t}(u) - \psi_\sigma(u)\ ,$$

$$J_n^+(t) = c_n^{-1/2} \sum_{i=1}^{n} W_{ni}(x, X_1, \ldots, X_n) H_t(Y_i - g(x)),$$
$$\tilde{J}_n^+(t) = J_n^+(t) - E(J_n^+(t))\ ,$$

$$J_n^-(t) = c_n^{-1/2} \sum_{i=1}^{n} W_{ni}(x, X_1, \ldots, X_n)\, I_t(Y_i - g(x))\ ,$$
$$\tilde{J}_n^-(t) = J_n^-(t) - E(J_n^-(t))\ .$$

The proof will be complete if we show

$$\lim_n \lim\ \sup_{d\to 0} P(\sup_{0\le t\le d} |\tilde{J}_n^+(t)| > \epsilon) = 0 \tag{3.2}$$

$$\lim_n \lim\ \sup_{d\to 0} \sup_{0\le t\le d} |E\, J_n^+(t)| = 0 \tag{3.3}$$

and the analogous result with $J_n^+(t)$ replaced by $J_n^-(t)$.

(3.3) follows easily from H1 and the dominated convergence theorem.

In order to show (3.2) it is enough to prove that the sequence $\widetilde{J}_n^+(t)$ of random variables on the space $C[0,1]$ of continuous function on $[0,1]$ is tight. According to Theorem 12.3 of Billingsley (1968) it suffices to verify that:

(i) the sequence $\{\widetilde{J}_n^+(0)\}$ is tight,
(ii) there exist constants $\gamma > 0$ and $\alpha > 1$ and a nondecreasing continuous function F on $[0,1]$ such that

$$E(|\widetilde{J}_n^+(t_2) - \widetilde{J}_n^+(t_1)|^\gamma) \leq (F(t_2) - F(t_1))^\alpha$$

holds for all $0 \leq t_1 < t_2$ and n large enough.

Since $\widetilde{J}_n^+(0) = 0$, (i) follows

$$E((\widetilde{J}_n^+(t_2) - \widetilde{J}_n^+(t_1))^2) \leq c_n^{-1} E[\sum_{i=1}^n W_{ni}^2(x)(H_{t_2}(Y_i - g(x)) - H_{t_1}(Y_i - g(x)))^2]$$

As in Lemma A of Fraiman (1980), we have that for $t_2 > t_1$

$$|H_{t_2}(u) - H_{t_1}(u)| \leq F(t_2) - F(t_1)$$

where $F(t) = |b|t - c^*(\sigma + t)^{-1}$, c^* is a positive constant and b is given in H1; which implies that $E(\widetilde{J}_n^+(t_2) - \widetilde{J}_n^+(t_1))^2 \leq c(F(t_2) - F(t_1))^2$.

Finally, a similar argument shows that the same result hold for $J_n^-(t)$.

Remark 3.1. If H4 holds and the kernel $K : R^d \to R$ is bounded and verifies that $\int |K(u)|du < \infty$, $\|u\|^d K(u) \to 0$ as $\|u\| \to \infty$ the related kernel weights and nearest neighbor with kernel weights satisfy that $c_n^{-1} E(\sum_{i=1}^n W_{ni}^2(x, X_1, \ldots, X_n))$ is bounded with $c_n^{-1} = nh_n^d$ or $c_n^{-1} = k_n$ respectively.

If H2 holds this condition is verified and so we may also apply Lemma 3.1 to the nearest neighbor weight.

Theorem 3.1. *Given $x_1, \ldots, x_p \in R^d$ verifying H10, the assumptions of Lemma 3.1 and such that $g_n(x_i) \to g(x_i)$, $s_n(x_i) \to s(x_i)$ $1 \leq i \leq p$ in probability, we have that $\{c_n^{-1/2}(g_n(x_i) - g(x_i))\ 1 \leq i \leq p\}$ has the same asymptotic distribution as*

$$\{s(x_i)[\widehat{\lambda}(x_i, g(x_i), s(x_i))]^{-1} c_n^{-1/2} \sum_{j=1}^n W_{nj}(x_i) Z_j^i \quad 1 \leq i \leq p\}$$

where $\widehat{\lambda}(x, t, \sigma) = \int \psi'((y - t)/\sigma) dF(y|X = x)$ and $Z_j^i = \psi(Y_j - g(x_i))/s(x_i))$.

We will now derive an explicit form for the asymptotic distribution of the robust nonparametric regression estimates related to kernel nearest neighbor and nearest neighbor with kernel weights, under the following assumptions:

N1. The kernel $K : R^d \to R$ is bounded, nonnegative, $0 < \int K(u)du < \infty$ and $\|u\|^d K(u) \to 0$ as $\|u\| \to \infty$.
N2. There exists $0 \leq \beta < \infty$ such that $h_n n^{1/(d+2)} \to \beta$ as $n \to \infty$.

N3. There exists a continuous, symmetric distribution function F_0 such that the conditional distribution $F(y|X = u) = F_0((y - g(u))/s(u))$ with g and s satisfying for $x = x_1, \dots, x_p$:

a) g verifies a Lipschitz condition of order one, and there exist $\lim_{\epsilon \to 0}(g(x + \epsilon u) - g(x))/\epsilon = g'(x, u)$.

b) s verifies a Lipschitz condition of order 1/2, i.e., $|s(u) - s(x)| < c\|u - x\|^{1/2}$ for some $c > 0$, and $\lim_{\epsilon \to 0}(s(x + \epsilon u) - s(x))/\epsilon^{1/2} = 0$.

Note that without loss of generality we may assume that the scale function of $F(y|X = u)$ is $s(u) = MAD_c(u)$.

N4. The kernel K is twice continuously differentiable and verifies:

a) $0 < \int |K_1(u)|du < \infty$; $\int K_1^2(u)du < \infty$ and $\|u\|^d K_1(u) \to 0$ as $\|u\| \to \infty$ where $K_1(u) = \sum_{j=1}^d \frac{\partial K}{\partial u_j}(u)u_j$.

b) $\|u\|^{p+1} K_2(u) \to 0$ as $\|u\| \to \infty$ where $K_2(u) = \sum_{i,j} \frac{\partial^2 K}{\partial u_i \partial u_j}(u)u_i u_j$ and $u = (u_1, \dots, u_d)$.

N5. There exists $0 \leq \beta \leq \infty$ such that $k_n^{\frac{1}{d}} n^{\frac{1}{d+2} - \frac{1}{d}} \to \beta$.

Let $v_{n1} \geq \dots \geq v_{nn} \geq 0$, $\sum_{i=1}^n v_{ni} = 1$ denote by $c_n = \sum_{i=1}^n v_{ni}^2$.

N6. $\lim_{n \to \infty} v_{n1} = 0$ and there exists a sequence of positive integers k_n such that $k_n \to \infty$, $k_n/n \to 0$ as $n \to \infty$ and $k_n v_{n1}$ is bounded and $\sum_{j > k_n} v_{nj} \to 0$ as $n \to \infty$.

N7. $v_{n1}/c_n^{1/2} \to 0$ as $n \to \infty$.

N8. $\lim_{n \to \infty} c_n^{-1} \sum_{j > k_n} v_{nj} = 0$ and $\lim_{n \to \infty} c_n^{-1/2}(k_n/n)^{1/d} = 0$.

Asymptotic distribution for kernel estimates.

Proposition 3.1. *Let $x_1, \dots, x_p$ be points in R^d verifying H4, H9, H10 and N3, and assume H1, N1, N2 and that Ψ is odd. Then, if $s_n(x_i) \to s(x_i)$ in probability for $1 \leq i \leq p$, we have that $\{(n h_n^d)^{1/2}(g_n(x_i) - g(x_i))\}_{i=1}^p$ converges in distribution to a vector of independent normal variables, the ith element of this vector having mean $b_{1,i} = \beta^{d/2+1} \int g'(x_i, u) K(u)\, du/(\int K(u)\, du)$ and variance $\sigma_{1,i}^2 \int \Psi^2(u)dF_0(u)/(\int \Psi'(u)dF_0(u))^2$ with $\sigma_{1,i}^2 = s^2(x_i) \int K^2(u)\, du/[f(x_i)(\int K(u)\, du)^2]$,where $g_n(x)$ is given in (1.2) with $W_{ni}(x) = K((X_i - x)/h_n)/\sum_{j=1}^n K((X_j - x)/h_n)$.*

Note that the asymptotic bias for the robust estimate is the same as for the linear kernel estimates of the regression function. A sufficient condition for the convergence of $s_n(x)$ is that the set $F_0^{-1}(1/4)$ is a single point.

In order to prove Proposition 3.1 we will use the following Lemma which is an easy modification of Theorem 2 of Schuster (1972) and which gives the asymptotic distribution of the Nadaraya-Watson estimates.

Lemma 3.2. *Let (X_i, Z_i) $1 \leq i \leq n$ be i.i.d. random vectors $X_i \in R^d$, $Z_i \in R^p$ with $|Z_i^j| \leq M$ for $1 \leq j \leq p$, $Z_i = (Z_i^1, \dots, Z_i^p)$. Denote by $F^{(j)}(z|X = u)$ the conditional distribution of $Z_1^j | X_1 = u$ and by $r_j(u) = E(Z_1^j | X_1 = u)$, $\sigma_j^2(u) = E((Z_1^j - r_j(x_j))^2 | X_1 = u)$, $\ell_{t_j}(u) = E(Z_1^j\, Z_1^t | X_1 = u)$. Let us suppose that:*

a) at each x_i, $1 \leq i \leq p$, r_j *is Lipschitz and*

$$\lim_{\epsilon \to 0}(r_j(x_i + \epsilon u) - r_j(x_i))/\epsilon = r_j'(x_i, u)$$

b) σ_j^2 *and* ℓ_{jt} *are continuous in a neighborhood of* x_i, $1 \leq i \leq p$.

Assume that $x_1, \ldots, x_p$ *verify H4, the kernel* K *verifies N1 and that we are given sequences* $(h_{in})_{n \in N}$ $1 \leq i \leq p$ *verifying N2 with* $\beta = \beta_i$, *for* $1 \leq i \leq p$.

Then $\{(n\, h_{in}^d)^{1/2}(r_n^i(x_i) - r_i(x_i))\}_{i=1}^p$ *is asymptotically distributed as a vector of independent normal variables, the ith variable having mean*

$$b_{2,i} = \beta_i^{\frac{d}{2}+1} \int r_i'(x_i, u)\, K(u)\, du / \int K(u)\, du \quad \text{and}$$

variance $\sigma_{2,i}^2 = \sigma_i^2(x_i) \int K^2(u)du/[f(x_i)(\int K(u)du)^2]$ *where* $r_n^i(x_i) = \sum_{j=1}^n Z_j^i K((X_j - x_i)/h_{in}) / \sum_{k=1}^n K((X_k - x_i)/h_{in})$.

Asymptotic distribution of nearest neighbor with kernel estimates.

Proposition 3.2. *Let* $x_1 \ldots x_p$ *be points in* R^d *verifying H4, H9, H10 and N3 and assume H1, N1, N4, N5 and that* Ψ *is odd.*

Denote by $g_n(x)$ *the solution of (1.2) with* $W_{ni}(x) = K((X_i - x)/H_n)/ \sum_{j=1}^n K((X_j - x)/H_n)$ $H_n = H_n(x)$. *Then if* $s_n(x_i) \to s(x_i)$ *in probability for* $1 \leq i \leq p$ *we have that* $\{k_n^{1/2}(g_n(x_i) - g(x_i))\}_{i=1}^p$ *converges in distribution to a vector of independent normal variables with the ith variable having mean* $b_{1,i}(f(x_i)\lambda(V_i))^{1/2}$ *and variance* $\sigma_{1,i}^2 f(x_i)\lambda(V_1) \int \Psi^2(u)dF_0(u)/(\int \Psi'(u)dF_0(u))^2$
where $\lambda(V_1)$ *is the Lebesgue measure of the unit ball and* $b_{1,i}$ *and* $\sigma_{1,i}^2$ *are given in Proposition 3.1.*

The proof of Proposition 3.2 will be obtained from the following Lemma which gives the asymptotic distribution of the classical nearest neighbor with kernel regression estimate.

Lemma 3.3. *Let* (X_i, Z_i) $1 \leq i \leq n$, *be i.i.d. random vectors* $X_i \in R^d$ $Z_i = (Z_i^1, \ldots, Z_i^p) \in R^p$, $|Z_i^j| \leq M$. *Let* $r_i(u)$, $\sigma_i^2(u)$ *and* $\ell_{ij}(u)$ *be as in Lemma 3.2. Assume that* $x_1 \ldots x_p$ *verify H4, and that N1, N4 and N5 are fulfilled, then we have that* $\{k_n^{1/2}(\widehat{r}_n^i(x_i) - r_i(x_i))\}_{i=1}^p$ *is asymptotically distributed as a vector of independent normal variables with the ith coordinate having mean* $b_{2,i}[f(x_i)\lambda(V_1)]^{1/2}$ *and variance* $\sigma_{2,i}^2 f(x_i)\lambda(V_1)$ *where* $b_{2,i}$ *and* $\sigma_{2,i}^2$ *are given in Lemma 3.2 with* $\beta_i = \beta$ *for all* i *and* $\widehat{r}_n^i(x_i) = \sum_{j=1}^n Z_j^i K((X_j - x_i)/H_n(x_i)) / \sum_{j=1}^n K((X_j - x_i)/H_n(x_i))$.

Asymptotic distribution for nearest weights.

Proposition 3.3. *Let* $x_1, \ldots, x_p$ *be points in* R^d *verifying H9, H10, N3 and assume H1, N6, N7, N8 and that* Ψ *is odd. Denote by* $g_n(x)$ *the solution of (1.2) with* $W_{nR_i}(x) = v_{ni}$. *Then if* $s_n(x_i) \to s(x_i)$ *in probability for* $1 \leq i \leq p$, *we have that* $\{c_n^{1/2}(g_n(x_i) - g(x_i)\}_{i=1}^p$ *converges in distribution to a normal zero mean vector with independent coordinates, the ith variable having variance* $s^2(x_i) \int \Psi^2(u)dF_0(u)/(\int \Psi'(u)dF_0(u))^2$. ▌

The proof of Proposition 3.3 can be obtained in a similar way as that of Proposition 3.1 using the consistency of $g_n(x)$ obtained in Boente and Fraiman (1989) and the following Lemma.

Lemma 3.4. *Let* (X_i, Z_i), $1 \le i \le n$ *be i.i.d. random vectors* $X_i \in R^d$ $Z_i = (Z_i^1, \dots, Z_i^p) \in R^p$ $|Z_i^j| \le M$. *Let* $r_i(u), \sigma_i^2(u)$ *and* $\ell_{ij}(u)$ *be as in Lemma 3.2 and* $r_n^i(x_i) = \sum_{j=1}^n W_{nj}(x_i) Z_j^i$ *with* $W_{nR_j}(x_i) = v_{nj}$ $R_j = R_j(x_i)$. *Then under N6, N7 and N8 we have that* $\{c_n^{-1/2}(r_n^i(x_i) - r_i(x_i))\}_{i=1}^p$ *is asymptotically distributed as a vector of independent normal variables zero mean and the ith variable has variance* $\sigma_i^2(x_i)$.

4. Proofs

Proof of Lemma 2.1. Denote by $\mathcal{A}_n = \{\|X_{R_{\ell_n}} - x\| < \delta\}$ with $\ell_n = [c/c_n]$. As $\ell_n \to \infty$, $\ell_n/n \to 0$, by Lemma 4 of Devroye (1982) we have $\sum_{n=1}^{\infty} P(\mathcal{A}_n^c) < \infty$. Therefore, in order to prove Lemma 2.1 it is enough to show that there exists $A_0 > 0$ such that

$$\sum_{n=1}^{\infty} P(\theta_n^{-1} \sup_y |F_n(y|X = x) - F(y|X = x) > A_0 \cap \mathcal{A}_n) < \infty .$$

Let $M = [\theta_n^{-1}]$, by H6 we may choose $y_k \in R$ $1 \le k \le m$ such that $(k-1)/M \le F(y_k|X = x) < k/M$. Thus

$$P(\theta_n^{-1} \sup_y |F_n(y|X = x) - F(y|X = x)| > A_0 \cap \mathcal{A}_n)$$
$$\le P(\theta_n^{-1} \max_{l \le k \le M} |F_n(y_k|X = x) - F(y_k|X = x)| > (A_0 - 1) \cap \mathcal{A}_n) .$$

Denote by $r_k(x) = F(y_k|X = x)$. By H6 we have $|r_k(X_{R_i}) - r_k(x)| < c\|X_{R_i} - x\|$ for $1 \le i \le \ell_n$ in $\mathcal{A}_n$ and therefore as

$$F_n(y_k|X = x) - F(y_k|X = x) = \sum_{i=1}^n W_{ni}(x)(I_{(-\infty, y_k]}(Y_i) - r_k(X_i))$$
$$+ \sum_{i=1}^n W_{ni}(x)(r_k(X_i) - r_k(x))$$

by H7 it suffices to show that there exist $A_1 > 0$ and $A_2 > 0$ such that

$$\sum_{n=1}^{\infty} P(\theta_n^{-1} \|X_{R_{\ell_n}} - x\| > A_1) < \infty \tag{4.1}$$

$$\sum_{n=1}^{\infty} \theta_n^{-1} \max_{1 \le k \le M} P(\theta_n^{-1} | \sum_{i=1}^n W_{ni}(x)(I_{(-\infty, y_k]}(Y_i) - r_k(X_i))| > A_2) < \infty \tag{4.2}$$

From Wagner (1973), $\widehat{f}_n(x) = \ell_n/(n\|X_{R_{\ell_n}} - x\|^d\lambda(V_1) \to f(x)$ completely as $n \to \infty$ where $\lambda(V_1)$ stands for the Lebesgue measure of the unit ball. By H8 $(\ell_n/n)^{1/d}\theta_n^{-1}$ is bounded and therefore (4.1) follows.

Finally, from Bernstein's inequality we have

$$M \max_{1\le k\le M} P(\theta_n^{-1}|\sum_{i=1}^{n} W_{ni}(x)(I_{(-\infty,y_k]}(Y_i) - r_k(X_i)| > A_2)$$

$$\le 2\theta_n^{-1}\exp(-A_2^2\theta_n^2/4c_n) \le 2\theta_n^{-1}n^{-A_2^2/4} \le 2a_0^{-1}c_n^{1/d}n^{-(A_2^2/4-1/d)} .$$

As $c_n \to 0$, (4.2) follows for A_2 large enough.

Proof of Theorem 3.1.

Denote by $\lambda_n(x,t,\sigma) = \sum_{i=1}^n W_{ni}(x)\Psi(y_i - t)/\sigma)$ and by $\widehat{\lambda}_n(x,t,\sigma) = \sum_{i=1}^n W_{ni}(x)$ $\Psi'((Y_i - t)/\sigma)$. The mean value theorem entails

$$0 = c_n^{-1/2}\lambda_n(x, g_n(x), s_n(x))s_n(x) = c_n^{-1/2}\lambda_n(x, g(x), s_n(x))s_n(x) -$$

$$c_n^{-1/2}(g_n(x) - g(x))\widehat{\lambda}_n(x, \xi_n(x), s_n(x)) \tag{4.3}$$

where $\xi_n(x) = (1-\theta_n)g_n(x) + \theta_n g(x)$ with $0 < \theta_n = \theta_n(x) < 1$.

As ψ' is bounded and uniformly continuous we have $\widehat{\lambda}(x_i, \xi_n(x_i), s_n(x_i)) \to \widehat{\lambda}(x_i, g(x_i), s(x_i))$ in probability, for all $1 \le i \le p$.

Lemma 3.1 of Boente and Fraiman (1989), implies that $\widehat{\lambda}_n(x_i, \xi_n(x_i), s_n(x_i)) - \widehat{\lambda}(x_i, \xi_n(x_i), s_n(x_i)) \to 0$ a.s. for $1 \le i \le p$ and therefore $\widehat{\lambda}_n(x_i, \xi_n(x_i), s_n(x_i)) \to \widehat{\lambda}(x_i, g(x_i), s(x_i))$ in probability. Finally, the conclusion of Theorem 3.1 follows from (4.3) and Lemma 3.1.

Proof of Proposition 3.1. From N1, N2, N3, N4 and Theorem 2.1 of Boente and Fraiman (1989) we obtain that $g_n(x_i)$ converges to $g(x_i)$ almost everywhere for $1 \le i \le p$. Applying Lemma 3.2 to $Z_j^i = \Psi((Y_j - g(x_i))/s(x_i))$ we obtain that $(nh_n^d)^{1/2}\sum_{j=1}^n W_{nj}(x)Z_j^i$ converges to a vector of independent normal variables, the ith-variable having mean $b_{1i}\int \Psi'(u)dF_0(u)/s(x_i)$ and variance $\sigma_{1,i}^2 \int \Psi^2(u)dF_0(u)/s^2(x_i)$ since $r_i(x_i) = 0$ $\sigma_i^2(x_i) = \int \Psi^2(u)dF_0(u)$ and $r_i'(x_i,u) = g'(x_i,u)\int \Psi'(t)dF_0(t)/s(x_i)$. Remark 3.1 and Lemma 3.1 complete the proof.

Proof of Lemma 3.3. Let $h_{in}^d = k_n/(n\,f(x_i)\,\lambda(V_1))$. Loftsgaarden and Quesenberry (1965) showed that $h_{in}^d/H_n^d(x_i) \to 1$ in probability; Moore and Yackel (1977) established that

$$(nH_n^d(x_i))^{-1}\sum_{j=1}^{n} K((X_j - x_i)/H_n(x_i) \to f(x_i)\int K(u)du$$

in probability, therefore as:

$$(k_n^{1/2}(\widehat{r}_n^1(x_1) - r_1(x_1)), \ldots, k_n^{1/2}(\widehat{r}_n^p(x_p) - r_p(x_p))$$

$$= (k_n^{1/2}(nh_{1n}^d)^{-1}\sum_{j=1}^{n}(Z_j^1 - r_1(x_1))\,K((X_j - x_1)/H_n(x_1)), \ldots,$$

$$k_n^{1/2}(nh_{pn}^d)^{-1}\sum_{j=1}^{n}(Z_j^p - r_p(x_p))\,K((X_j - x_p)/H_n(x_p)))\,\mathrm{diag}\,(\lambda_1^n, \ldots, \lambda_p^n)$$

with $(\lambda_i^n)^{-1} = (H_n(x_i)/h_{in})^d (n\, H_n^d(x_i))^{-1} \sum_{j=1}^n K((x_j - x_i)/H_n(x_i))$ it is enough to see that

$$S_n = \{k_n^{1/2}(n\, h_{in}^d)^{-1} \sum_{j=1}^n (Z_j^i - r_i(x_i))\, K((X_j - x_i)/H_n(x_i))\}_{i=1}^p$$

is asymptotically normally distributed with covariances 0 and the ith coordinate with mean

$$\widetilde{b}_i = b_{2,i}(f(x_i)\lambda(V_1))^{-1/d} f(x_i) \int K(u)\, du$$

and variance $\widetilde{\sigma}_i^2 = \sigma_{2,i}^2 f(x_i)\lambda(V_1) f^2(x_i)(\int K(u)\, du)^2$.

A second order Taylor expansion gives $S_n = S_{n1} + S_{n2} + S_{n3}$ with $S_{nj} = (S_{nj}^{(1)}, \ldots, S_{nj}^{(p)})$ and

$$S_{n1}^{(i)} = k_n^{1/2}(n\, h_{in}^d)^{-1} \sum_{j=1}^n K((X_j - x_i)/h_{in})(Z_j^i - r_i(x_i))$$

$$S_{n2}^{(i)} = k_n^{1/2}\, V_{in}(n\, h_{in}^d)^{-1} \sum_{j=1}^n K_1((X_j - x_i)/h_{in}(Z_j^i - r_i(x_i))$$

$$S_{n3}^{(i)} = k_n^{1/2}\, V_{in}^2(n\, h_{in}^d)^{-1} \sum_{j=1}^n K_2((X_j - x_i)/\xi_n^{(i)})(Z_j^i - r_i(x_i))$$

where $\min(h_{in}, H_n(x_i)) \leq \xi_n^{(i)} \leq \max(h_{in}, H_n(x_i))$, and $V_{in} = (h_{in}/H_n(x_i)) - 1$.

By Lemma 3.2 $S_{n1} \xrightarrow{w} N(\widetilde{b}, \sum)$ with $\widetilde{b} = (\widetilde{b}_1, \ldots, \widetilde{b}_p)$ and $\sum = \operatorname{diag}(\widetilde{\sigma}_1^2, \ldots, \widetilde{\sigma}_p^2)$, and $\xrightarrow{w}$ stands for weak convergence. Therefore it is enough to show that S_{n2} and S_{n3} converge to 0 in probability.

Moore and Yackel (1977) have shown that $k_n^{1/2}((h_{in}/H_n(x_i))^d - 1)$ is asymptotically normally distributed, thus it suffices to see that, for all $1 \leq i \leq p$:

$$\text{(a)} \quad (n\, h_{in}^d)^{-1} \sum_{j=1}^n K_1((X_j - x_i)/h_{in})(Z_j^i - r_i(x_i)) \to 0 \text{ in probability}$$

and

$$\text{(b)} \quad (n\, h_{in}^d)^{-1} \sum_{j=1}^n K_2((X_j - x_i)/\xi_n^i)(Z_j^i - r_i(x_i)) \text{ is bounded in probability.}$$

By Tchebischev's inequality (a) follows easily from Bochner's theorem since r_i and σ_i^2 are continuous at x_i .

(b) may be obtained from N4(b) in a similar way as in Theorem 5 from Boente and Fraiman (1991) since Z_j^i are bounded.

Proof of Proposition 3.2. If $K(uz) \geq K(z)$ for $z \in R^d$ $u \in [0,1]$ the convergence of $g_n(x_i)$ to $g(x_i)$ follows from Collomb (1980). If K verifies (2.1) this result follows from Theorem 3.1 of Boente and Fraiman (1989). If not, a direct proof can be given using Theorem 2.2 of Boente and Fraiman (1989) and establishing that $F_n(y|X=x_i) \to F(y|X=x_i)$ a.s. by a second order Taylor's expansion as in Lemma 3.3.

Now the proof follows as in Proposition 3.1 using Lemma 3.3.

Proof of Lemma 3.4. Denote by S_{1n} and S_{2n} the vectors with coordinates

$$S_{1n}^{(i)} = \sum_{j=1}^{n} W_{nj}(x_i)(Z_j^i - r_i(X_j))$$

$$S_{2n}^{(i)} = \sum_{j=1}^{n} W_{nj}(x_i)(r_i(X_j) - r_i(x_i))$$

We will show that

(a) $c_n^{-1/2}\, S_{1n} \xrightarrow{w} N(0,\sum)$ $\sum = \text{diag}\,(\sigma_1^2(x_1),\ldots,\sigma_p^2(x_p))$

(b) $c_n^{-1/2}\, S_{2n} \to 0$ in probability;

which entails the Lemma.

We begin by proving (b).

$$\begin{aligned} c_n^{-1/2}\, S_{2n}^{(i)} &= c_n^{-1/2} \sum_{j=1}^{k_n} v_{nj}(r_i(X_{R_j(x_i)}) - r_i(x_i)) \\ &\quad + c_n^{-1/2} \sum_{j>k_n} v_{nj}(r_i(X_{R_j(x_i)}) - r_i(x_i))\,. \end{aligned} \tag{4.4}$$

As $|r_i(x)| \le M$ for all i, x, N8 implies that the second term on the right side of (4.4) converges to 0. As r_i is Lipschitz, the first term can be majorized by $C\, c_n^{-1/2}\|X_{R_{k_n}(x_i)} - x_i\| = C\, c_n^{-1/2} H_n(x_i)$ with C a positive constant. Thus the convergence in probability of $k_n/(n\, H_n^p(x_i))$ to $f(x_i)\,\lambda(V_1)$ and N8 concludes the proof of (b).

In order to prove (a) let $a \in R^p$, $\|a\| = 1$, we have to show that $a'\, c_n^{-1/2} \sum^{-1/2} S_{1n} \xrightarrow{w} N(0,1)$ where $\sum^{-1/2} = \text{diag}\,((\sigma_i^2(x_i))^{-1/2})$.

Denote by Λ_n the random diagonal matrix in $R^{p\times p}$ with elements $\gamma_i(X) = c_n^{-1} \sum_{j=1}^{n} W_{nj}^2(x_i)[\sigma_i^2(X_j) - (r_i(x_i) - r_i(X_j))^2]$. We will show that:

(i) $\Lambda_n \to \sum$ in probability

(ii) $a'\, c_n^{-1/2}\ \Lambda_n^{-1/2}\ S_{1n} \xrightarrow{w} N(0,1)$

Since

$$\begin{aligned} \gamma_i(X) =& \sigma_i^2(x_i) + c_n^{-1} \sum_{j=1}^{n} W_{nj}^2(x_i)(\sigma_i^2(X_j) - \sigma_i^2(x_i)) \\ & - c_n^{-1} \sum_{j=1}^{n} W_{nj}^2(x_i)(r_i(x_i) - r_i(X_j))^2 \end{aligned}$$

(i) follows if the second and third term converge to zero in probability, but this follows easily from the continuity of σ_i and r_i, from assumption N8 and the fact that $H_n(x_i) \to 0$ in probability.

We will derive (ii) from the Berry-Essen Theorem. For each fixed i, $1 \leq i \leq p$, let V_j^i be i.i.d. random variables with the distribution of $Z_j^i|X_j = \xi_j$ and $U_{jn} = c_n^{-1/2} \sum_{i=1}^p a_i \gamma_i^{-1/2}(\xi) W_{nj}(x_i,\xi)(V_j^i - E(V_j^i))$ then $\mathcal{L}(a' c_n^{-1/2} \Lambda_n^{-1/2} S_{1n}|X = \xi) = \mathcal{L}(\sum_{j=1}^n U_{jn})$ where $\mathcal{L}(U)$ denotes the distribution of U.

We have that U_{jn} are independent zero mean random variables for $1 \leq j \leq n$ such that

$$\sum_{j=1}^n \text{Var}\,(U_{jn}) = \sum_{j=1}^n c_n^{-1} \sum_{i=1}^p a_i^2 \gamma_i^{-1}(\xi) W_{jn}^2(x_i,\xi)\text{Var}(V_j^i) + c_n$$
$$= \sum_{i=1}^p a_i^2 \gamma_i^{-1}(\xi) c_n^{-1} \sum_{j=1}^n W_{nj}^2(x_i,\xi) E[(Z_j^i - r_i(\xi_j))^2|X_j = \xi_j] + c_n$$
$$= \sum_{i=1}^p a_i^2 + c_n \to 1$$

since by N6 and N8, the covariance term c_n converges to 0. On the other hand, $|U_{jn}| \leq M\, c_n^{-1/2}\, v_{1n} \sum_{i=1}^p |a_i| \gamma_i^{-1/2}(\xi) = M_n(\xi)$. Then $\Gamma_n = \sum_{j=1}^n E|U_{jn}|^3 \leq 2M_n(\xi)$ and the Berry-Essen theorem apply to U_{jn} since $c_n^{-1/2} v_{n1} \to 0$ as $n \to \infty$. Thus, there is a universal constant A_0 such that

$$|P(a' c_n^{-1/2}\, \Lambda_n^{-1/2}\, S_{1n} \leq z|X = \xi) - \phi(z)| \leq A_0\, \Gamma_n \leq A_0\, M_n(\xi)$$

Therefore

$$|P(a' c_n^{-1/2}\, \Lambda_n^{-1/2}\, S_{n1n} \leq z) - \phi(z)|$$
$$\leq A_0\, M\, c_n^{-1/2}\, v_{n1} A\, P(\sum_{i=1}^p |a_i| \gamma_i(X)^{-1/2} < A)$$
$$+ 2P(\sum_{i=1}^p |a_i| \gamma_i(X)^{-1/2} > A)$$

which establishes (ii) since $\Lambda_n \to \sum$ in probability which is positive definite and $c_n^{-1/2} v_{n1} \to 0$ as $n \to \infty$.

References

BILLINGSLEY, P., *Convergence of probability measures.* J. Wiley, New York, 1968

BOENTE, G., FRAIMAN, R., *Robust nonparametric regression.* J. Multivariate Anal. 29, 180-198, 1989

BOENTE, G., FRAIMAN, R., *Strong order of convergence and asymptotic distribution of nearest neighbor density estimates from dependent observations.* Trabajos de Matemática No. 116, I.A.M. CONICET. To appear in Sankhya, Serie A, Vol. 53., 1991

COLLOMB, G., *Estimation de la regression par la méthode des k points les plus proches avec noyau: Quelques propiétés de convergence ponctuelle.* Lecture Notes in Mathematics, 821, 159-175, 1980.

COVER, T.M., *Estimation by the nearest neighbor rule.* IEEE Trans. Inform. Theory 14, 50-55, 1968.

DEVROYE, L., *The uniform convergence of nearest neighbor regression function estimators and their application in optimization.* IEEE Trans. Inform. Theory 24, 142-151, 1978.

DEVROYE, L., *On the almost everywhere convergence of nonparametric regression function estimates.* Ann. Statist. 9, 1310-1319, 1981.

DEVROYE, L., *Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates.* Z. Wahrsch. Verw. Gebiete 61, 467-481, 1982.

DEVROYE, L., WAGNER, T. J., *On the L^1 convergence of kernel estimators of regression functions with applications in discrimination.* Z. Wahrsch. Verw. Gebiete 51, 15-25, 1980 a.

DEVROYE, L., WAGNER, T. J., *Distribution-free consistency results in nonparametric discrimination and regression function estimation.* Ann. Statist. 8, 231-239, 1980 b.

FRAIMAN, R., *Estimadores robustos para regresión no lineal.* Doctoral dissertation, Universidad de Buenos Aires, 1980.

GREBLICKI, W., KRZYZAK, A. and PAWLAK, M., *Distribution-free pointwise consistency of kernel regression estimate.* Ann. Statist. 12, 1570-1575, 1984.

GYORFI, L. *Recent results on nonparametric regression estimate and multiple classification.* Problems Control Inform. Theory 10, 43-52, 1981.

LOFTSGAARDEN, D.O., QUESENBERRY, C.D., *A nonparametric estimate for a multivariate density function.* Ann. Math. Statist. 36, 1049-1051, 1965.

MACK, Y.P., *Local properties of $k-NN$ regression estimates.* Siam J. Alg. Disc. Meth. 2, 311-323, 1981.

MACK, Y.P., ROSENBLATT, M., *Multivariate k-nearest neighbor density estimates.* J. of Mult. Anal. 9, 1-15, 1979.

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

NADARAYA, E.A., *On estimating regression.* Th. Prob. Appl. 9, 141-142, 1964.

SCHUSTER, E.F., *Joint asymptotic distribution of the estimated regression function at a finite number of distinct points.* Ann. Math. Statist. 43, 84-88, 1972.

SPIEGELMAN, C., SACKS, J. *Consistent window estimation in nonparametric regression.* Ann. Statist. 8, 240-246, 1980.

STONE, C., *Consistent nonparametric regression.* Ann. Statist. 5, 595-645, 1977.

WATSON, G.S., *Smooth regression analysis.* Sankhya, Series A 26, 359-372, 1964.

Universidad de Buenos Aires and
CONICET
Buenos Aires – Argentina.

Universidad de la República
Montevideo - Uruguay
and
Universidad de Buenos Aires
Buenos Aires - Argentina.

Recibido por UMA el 16 de mayo de 1989.