ESTIMATING THE ERROR DISTRIBUTION FUNCTION IN NONPARAMETRIC REGRESSION WITH MULTIVARIATE COVARIATES

URSULA U. MÜLLER, ANTON SCHICK AND WOLFGANG WEFELMEYER

Abstract. We consider nonparametric regression models with multivariate covariates and estimate the regression curve by an undersmoothed local polynomial smoother. The resulting residual-based empirical distribution function is shown to differ from the error-based empirical distribution function by the density times the average of the errors, up to a uniformly negligible remainder term. This result implies a functional central limit theorem for the residual-based empirical distribution function.

1. Introduction and Main Results

We consider the nonparametric regression model

$$Y = r(X) + \varepsilon,$$

where the error ε has mean zero, finite variance and is independent of the m-dimensional covariate vector X which is assumed to be *quasi-uniform* on the unit cube $\mathscr{C} = [0,1]^m$. By the latter we mean that X has a density g that is bounded and bounded away from zero on $\mathscr{C}$ and is zero otherwise. We are interested in estimating the error distribution function F by a residual-based empirical distribution function. This problem was already addressed by Müller, Schick and Wefelmeyer (2007) in the case $m = 1$. They used residuals based on an undersmoothed local linear smoother for the regression function. Here we follow their approach, but use local polynomial smoothers in order to cover multivariate covariates.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n independent copies of of (X, Y). In order to define the local polynomial smoother we introduce some notation. By a *multi-index* we mean an m-dimensional vector $i = (i_1, \dots, i_m)$ whose components are non-negative integers. For a multi-index i let ψ_i denote the function on $\mathbb{R}^m$ defined by

$$\psi_i(x) = \frac{x_1^{i_1}}{i_1!} \cdots \frac{x_m^{i_m}}{i_m!}, \quad x = (x_1, \dots, x_m) \in \mathbb{R}^m.$$

Set $i_\bullet = i_1 + \dots + i_m$. For a non-negative integer k, let $I(k)$ denote the set of multi-indices i with $i_\bullet \le k$ and $J(k)$ the set of multi-indices i with $i_\bullet = k$. Now fix densities $w_1, \dots, w_m$ and set

$$w(x) = w_1(x_1) \cdots w_m(x_m), \quad x = (x_1, \dots, x_m) \in \mathbb{R}^m.$$

Anton Schick was supported by NSF Grant DMS 0405791.

Let c_n be a bandwidth. Fix a non-negative integer d. Then the *local polynomial smoother* $\hat{r}$ *(of degree d)* is defined as follows. For an x in $\mathscr{C}$, $\hat{r}(x)$ is the component $\hat{\beta}_0$ corresponding to the multi-index $0 = (0, \dots, 0)$ of a minimizer

$$\hat{\beta} = \operatorname*{arg\,min}_{\beta=(\beta_i)_{i\in I(d)}} \sum_{j=1}^{n} \Big(Y_j - \sum_{i\in I(d)} \beta_i \psi_i\Big(\frac{X_j - x}{c_n}\Big)\Big)^2 w\Big(\frac{X_j - x}{c_n}\Big).$$

To state our main result we also need to introduce the Hölder spaces $H(k,\gamma)$ for $k = 0, 1, \dots$ and $0 < \gamma \le 1$. A function h from $\mathscr{C}$ to $\mathbb{R}$ belongs to $H(k,\gamma)$ if it has continuous partial derivatives up to order k and the partial derivatives of order k are Hölder with exponent γ. For such functions h we define the norm

$$\|h\|_{k,\gamma} = \max_{i\in I(k)} \sup_{x\in\mathscr{C}} |D^i h(x)| + \max_{i\in J(k)} \sup_{x,y\in\mathscr{C}, x\neq y} \frac{|D^i h(y) - D^i h(x)|}{\|x-y\|^\gamma}$$

where $\|v\|$ denotes the euclidean norm of a vector v and

$$D^i h(x) = \frac{\partial^{i_\bullet}}{\partial x_1^{i_1} \cdots \partial x_m^{i_m}} h(x), \quad x = (x_1, \dots, x_m) \in \mathscr{C}.$$

Let $H_1(k,\gamma)$ denote the unit ball of $H(k,\gamma)$ for this norm. The following result will be proved in the next section.

Lemma 1. *Suppose the regression function r belongs to $H(d,\gamma)$ with $s = d + \gamma > 3m/2$, the error variable has mean zero and a finite moment of order $\zeta > 4s/(2s-m)$, and the densities $w_1, \dots, w_m$ are $(m+2)$-times continuously differentiable and have compact support $[-1,1]$. Let $c_n \sim (n \log n)^{-1/(2s)}$. Then there is a random function $\hat{a}$ such that*

$$P(\hat{a} \in H_1(m,\alpha)) \to 1 \tag{1.1}$$

for some $\alpha > 0$,

$$\int |\hat{a}(x)|^{1+\xi} g(x)\, dx = o_p(n^{-1/2}) \tag{1.2}$$

for $\xi > m/(2s-m)$,

$$\int \hat{a}(x) g(x)\, dx = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}), \tag{1.3}$$

and

$$\sup_{x\in\mathscr{C}} |\hat{r}(x) - r(x) - \hat{a}(x)| = o_p(n^{-1/2}). \tag{1.4}$$

Remark 1. If r has continuous partial derivatives of order $s \ge 1$, then it belongs to $H(s-1,1)$ and the above lemma applies with $d = s-1$ and $\gamma = 1$ provided $s > 3m/2$. However, if $s > 3m/2$ and we choose the degree d to be s, then the conclusion of the lemma still holds if we take $c_n \sim n^{-1/(2s)}$. Indeed, inspecting the proof of the lemma shows that then the left-hand side of (2.4) is of order $o(\|y-x\|^s)$ which implies that the left-hand side of (2.8) is of order $o_p(c_n^s) = o_p(n^{-1/2})$.

Theorem 1. *Suppose that the assumptions of the previous lemma are met and the error variable has a density f that is Hölder with exponent $\xi > m/(2s-m)$. Then*

$$\sup_{t\in\mathbb{R}} \Big| \frac{1}{n}\sum_{j=1}^n \mathbf{1}[Y_j - \hat{r}(X_j) \le t] - \frac{1}{n}\sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t] - f(t)\frac{1}{n}\sum_{j=1}^n \varepsilon_j \Big| = o_p(n^{-1/2}).$$

Proof. By Corollary 2.7.2 in van der Vaart and Wellner (1996), there is a constant K such that

$$\log N_{[\,]}(\eta^2, H_1(m,\alpha), L_1(G)) \le K\eta^{-2m/(m+\alpha)}, \quad 0 < \eta. \tag{1.5}$$

Therefore

$$\int_0^1 \sqrt{\log N_{[\,]}(\eta^2, H_1(m,\alpha), L_1(G))}\, d\eta < \infty.$$

Consequently, Theorem 2.2 in Müller, Schick and Wefelmeyer (2007) applied with $\mathcal{D} = H_1(m,\alpha)$, boundedness of f, and the expansion (1.3) imply the desired result. □

Remark 2. For parametric regression, analogous results to Theorem 1 can be found in Koul (2002). In such models, the regression function can be estimated at the faster parametric rate of convergence, and one thus gets by with weaker assumptions on the error density. Uniform continuity of the error density and weaker moment conditions suffice. For example, for linear regression $r(x) = \beta^\top x$ one obtains the expansion

$$\sup_{t\in\mathbb{R}} \Big| \frac{1}{n}\sum_{j=1}^n \mathbf{1}[Y_j - \hat{\beta}^\top X_j \le t] - \frac{1}{n}\sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t] - f(t)E[X]^\top(\hat{\beta}-\beta) \Big| = o_p(n^{-1/2}),$$

with $\hat{\beta}$ the least squares estimator. This only requires that ε has mean zero, finite variance and a uniformly continuous density and that the matrix $E[XX^\top]$ is invertible. If one of the coordinates of X equals 1 or if more generally $c^\top X$ equals 1 almost surely for some vector c, then one has $E[X]^\top(E[XX^\top])^{-1}X = 1$ almost surely and therefore

$$E[X]^\top(\hat{\beta}-\beta) = E[X]^\top \frac{1}{n}\sum_{j=1}^n (E[XX^\top])^{-1}X_j\varepsilon_j + o_p(n^{-1/2}) = \frac{1}{n}\sum_{j=1}^n \varepsilon_j + o_p(n^{-1/2}),$$

and one consequently obtains the exact analogue of the expansion in Theorem 1. To see that $E[X]^\top(E[XX^\top])^{-1}X = 1$ almost surely we note that $E[X]^\top(E[XX^\top])^{-1}X$ equals $E[X]^\top A(E[A^\top XX^\top A])^{-1}A^\top X$ for any invertible $m\times m$ matrix A. Now take A such that its first column equals c and the other columns form a basis for $\{v \in \mathbb{R}^m : v^\top E[X] = 0\}$ and are orthogonal for the inner product $(v,w) = E[v^\top XX^\top w]$. Then $E[A^\top XX^\top A]$ equals the $m\times m$ identity matrix, $E[X]^\top A$ equals $(1,0,\dots,0)$, and $E[X]^\top A(E[A^\top XX^\top A])^{-1}A^\top X$ equals $c^\top X$.

Remark 3. The expansion in Theorem 1 implies that the residual-based empirical process

$$n^{-1/2}\sum_{j=1}^{n}(\mathbf{1}[Y_j-\hat r(X_j)\le t]-F(t)),\quad -\infty\le t\le\infty,$$

converges weakly in $D[-\infty,\infty]$ to a centered Gaussian process with covariance function

$$(s,t)\mapsto F(s\wedge t)-F(s)F(t)+f(s)c(t)+f(t)c(s)+f(s)f(t)\sigma^2,$$

where $c(t)=\int_{-\infty}^{t}xf(x)\,dx$ is the mean of $\varepsilon\mathbf{1}[\varepsilon_1\le t]$.

Remark 4. Note that $4s/(2s-m)<3$ and $m/(2s-m)<1/2$ for $s>3m/2$. Thus the assumptions on the error variable in the above theorem are met if the error variable has a finite moment of order 3 and if its density f is Hölder with exponent $1/2$. The latter is implied if f has finite Fisher information for location; see e.g. Koul (2002, page 79). Thus the conclusion of the above theorem holds if the error density f has mean zero, a finite moment of order 3 and finite Fisher information for location, and if the regression function r belongs to a Hölder space $H(d,\gamma)$ with $d+\gamma>3m/2$.

Remark 5. For $m=1$, and using a linear smoother, Müller, Schick and Wefelmeyer (2007) obtained the assertion of Theorem 1 assuming that r is twice continuously differentiable, the error distribution has a moment of order greater than $8/3$, and the error density is Hölder of order greater than $1/3$. Here we allow for higher degree smoothers. Thus we can relax the moment and smoothness assumptions on the error distribution at the expense of more smoothness on the regression function. For example, if the second derivative of r is Lipschitz and a quadratic smoother is used, then it suffices that F has a finite moment of order greater than $12/5$ and f is Hölder of order greater than $1/5$.

2. Proof of Lemma 1

Abbreviate $I(d)$ by I. By the choice of bandwidth and since $2s>3m$ we have

$$\rho_n=\frac{\log n}{nc_n^m}\sim\frac{(\log n)^{1+m/(2s)}}{n^{1-m/(2s)}}\le(\log n)^{4/3}n^{-2/3}.$$

The minimizer $\hat\beta$ must satisfy the normal equations

$$C_i(x)-\sum_{k\in I}\hat Q_{ik}(x)\hat\beta_k=0,\quad i\in I,$$

where

$$C_i(x)=\frac{1}{nc_n^m}\sum_{j=1}^{n}Y_j\psi_i\Big(\frac{X_j-x}{c_n}\Big)w\Big(\frac{X_j-x}{c_n}\Big)$$

and

$$\hat Q_{ik}(x)=\frac{1}{nc_n^m}\sum_{j=1}^{n}\psi_i\Big(\frac{X_j-x}{c_n}\Big)\psi_k\Big(\frac{X_j-x}{c_n}\Big)w\Big(\frac{X_j-x}{c_n}\Big).$$

Since $Y_j = r(X_j) + \varepsilon_j$, we can write $C_i(x) = A_i(x) + B_i(x)$ where

$$A_i(x) = \frac{1}{nc_n^m} \sum_{j=1}^{n} \varepsilon_j \psi_i\Big(\frac{X_j - x}{c_n}\Big) w\Big(\frac{X_j - x}{c_n}\Big)$$

and

$$B_i(x) = \frac{1}{nc_n^m} \sum_{j=1}^{n} r(X_j) \psi_i\Big(\frac{X_j - x}{c_n}\Big) w\Big(\frac{X_j - x}{c_n}\Big).$$

Let $Q_{ik}(x) = E[\hat{Q}_{ik}(x)]$. It follows from Corollary 1 applied with $T = 1$, $\beta = \infty$ and $v = \psi_i \psi_k w$ that

$$\sup_{x \in \mathscr{C}} |\hat{Q}_{ik}(x) - Q_{ik}(x)| = O_p(\rho_n^{1/2}) \tag{2.1}$$

and applied with $T = \varepsilon$, $\beta > 4s/(2s - m)$ and $v = D^i(\psi_k w)$ that

$$\sup_{x \in \mathscr{C}} |c_n^{i_\bullet} D^i A_k(x)| = O_p(\rho_n^{1/2}), \quad i \in I(m+1). \tag{2.2}$$

Note that

$$Q_{ik}(x) = \int \psi_i(u) \psi_k(u) g(x + c_n u) w(u)\, du.$$

Since X is quasi-uniform on $\mathscr{C}$, one now finds constants $0 < \lambda < \Lambda < \infty$ such that

$$\lambda < \sum_{i,k \in I} a_i Q_{ik}(x) a_k < \Lambda \tag{2.3}$$

for all $x \in \mathscr{C}$, all a_i, $i \in I$, with $\sum_{i \in I} a_i^2 = 1$ and all large n.

Since r belongs to $H(d, \gamma)$, we obtain

$$\Big| r(y) - \sum_{k \in I} D^k r(x) \psi_k(y - x) \Big| \le M \|y - x\|^{d+\gamma} \tag{2.4}$$

for some finite constant M and all x, y in $\mathscr{C}$. From this we see that

$$\Big| B_i(x) - \sum_{k \in I} \hat{Q}_{ik}(x) \bar{\beta}_k(x) \Big| \le M c_n^{d+\gamma} \hat{Q}_{00}(x), \quad x \in \mathscr{C}, \tag{2.5}$$

where $\bar{\beta}_k(x) = c_n^{k_\bullet} D^k r(x)$.

Let now $\hat{Q}(x)$ denote the matrix with entries $\hat{Q}_{ik}(x)$, $i, k \in I$, $Q(x)$ the matrix with entries $Q_{ik}(x)$, $i, k \in I$, $A(x)$ the vector with components $A_i(x)$ and $B(x)$ the vector with components $B_i(x)$. In view of (2.3) the eigen values of $Q(x)$ are in the interval $[\lambda, \Lambda]$. Thus $Q(x)$ has an inverse $Q^{-1}(x)$ with eigen values in $[1/\Lambda, 1/\lambda]$. Hence $\hat{Q}(x)$ is invertible on the event $\{\sup_{x \in \mathscr{C}} \|\hat{Q}(x) - Q(x)\| \le \lambda/2\}$ whose probability tends to one by (2.1). On this event we have

$$\hat{r}(x) = e^\top \hat{Q}^{-1}(x) C(x), \quad x \in \mathscr{C},$$

where $e = (e_i)_{i \in I}$ is such that $e_0 = 1$ and $e_i = 0$ for $i \neq 0$. Since $r(x) = e^\top \bar{\beta}(x)$, we obtain on this event the identity

$$\begin{aligned}\hat{r}(x) - r(x) &= e^\top \hat{Q}^{-1}(x)(C(x) - \hat{Q}(x)\bar{\beta}(x))\\ &= e^\top \hat{Q}^{-1}(x)A(x) + e^\top \hat{Q}^{-1}(x)[B(x) - \hat{Q}(x)\bar{\beta}(x)].\end{aligned}$$

It follows from (2.1), (2.2), (2.5) and the choice of c_n that

$$\sup_{x \in \mathscr{C}} \|\hat{Q}^{-1}(x) - Q^{-1}(x)\| = O_p(\rho_n^{1/2}), \tag{2.6}$$

$$\sup_{x \in \mathscr{C}} \|c_n^{i_\bullet} D^i A(x)\| = O_p(\rho_n^{1/2}), \quad i \in I(m+1) \tag{2.7}$$

$$\sup_{x \in \mathscr{C}} \|B(x) - \hat{Q}(x)\bar{\beta}(x)\| = O_p(c_n^{d+\gamma}) = o_p(n^{-1/2}). \tag{2.8}$$

Thus if we take

$$\hat{a}(x) = e^\top Q^{-1}(x)A(x), \quad x \in \mathscr{C},$$

we obtain the desired (1.4).

One verifies directly that $\sup_{x \in \mathscr{C}} \|c_n^{i_\bullet} D^i Q(x)\| = O(1)$. Note also that

$$\frac{\partial}{\partial x_k} Q^{-1}(x) = Q^{-1}(x)\Big[\frac{\partial}{\partial x_k} Q(x)\Big] Q^{-1}(x), \quad k = 1, \dots, m.$$

Thus we see that

$$\sup_{x \in \mathscr{C}} |D^i \hat{a}(x)| = O_p(c_n^{-i_\bullet} \rho_n^{1/2}), \quad i \in I(m+1).$$

This shows that

$$\max_{i \in I(m)} \sup_{x \in \mathscr{C}} |D^i \hat{a}(x)| = O_p(c_n^{-m} \rho_n^{1/2}) = o_p(n^{-b/2})$$

for some $0 < b < 1 - 3m/(2s)$. In view of $2s > 3m$, we have $(3m + 2 - 2s)/(4s) < 1/(2s)$ and thus

$$\max_{i \in J(m+1)} \sup_{x \in \mathscr{C}} |D^i \hat{a}(x)| = O_p(c_n^{-(m+1)} \rho_n^{1/2}) = o_p(n^{1/(2s)}).$$

With $\alpha = \min(b/2, 1 - 1/(2s))$ we obtain

$$\max_{i \in J(m)} \sup_{y \neq x, \|y - x\| > 1/n} \frac{|D^i \hat{a}(y) - D^i \hat{a}(x)|}{\|y - x\|^\alpha} = o_p(n^{\alpha - b/2}) = o_p(1)$$

and

$$\max_{i \in J(m)} \sup_{y \neq x, \|y - x\| \le 1/n} \frac{|D^i \hat{a}(y) - D^i \hat{a}(x)|}{\|y - x\|^\alpha} = o_p(n^{\alpha - 1 + 1/(2s)}) = o_p(1).$$

This establishes (1.1) as α is positive.

Next we have

$$\int |\hat{a}(x)|^{1+\xi} g(x)\, dx \le \sup_{x \in \mathscr{C}} \|\hat{a}(x)\|^{1+\xi} \le (1/\lambda)^{1+\xi} \sup_{x \in \mathscr{C}} \|A(x)\|^{1+\xi} = O_p(\rho_n^{(1+\xi)/2}).$$

We obtain the desired (1.2) as

$$\Big(1-\frac{m}{2s}\Big)(1+\xi) > \Big(1-\frac{m}{2s}\Big)\Big(1+\frac{m}{2s-m}\Big) = 1$$

by the choice of ξ.

We can write

$$\int \hat{a}(x)g(x)\,dx = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \Delta_n(X_j),$$

where

$$\begin{aligned}\Delta_n(X) &= c_n^{-m}\int e^\top Q^{-1}(x)\psi\Big(\frac{X-x}{c_n}\Big)w\Big(\frac{X-x}{c_n}\Big)g(x)\,dx \\ &= \int e^\top Q^{-1}(X-c_n u)\psi(u)w(u)g(X-c_n u)\,du\end{aligned}$$

with $\psi = (\psi_i)_{i\in I}$. Thus (1.3) follows if we show that $E[(\Delta_n(X)-1)^2] \to 0$. Since X is quasi-uniform and Δ_n is bounded by M/λ, where

$$M = \sup_{u\in[-1,1]^m} \|\psi(u)\| w(u) \sup_{x\in\mathscr{C}} g(x),$$

the desired $E[(\Delta_n(X)-1)^2] \to 0$ follows if we show

$$R_n = E[\mathbf{1}[X\in\mathscr{C}_n](\Delta_n(X)-1)^2] \to 0,$$

with $\mathscr{C}_n = [c_n, 1-c_n]^m$.

Let $Q_*(x) = g(x)\Psi$ where Ψ is the invertible matrix with entries

$$\Psi_{ik} = \int \psi_i(u)\psi_k(u)w(u)\,du, \quad i,k\in I.$$

Then we can write

$$1 = e^\top\Psi^{-1}\Psi e = e^\top\Psi^{-1}\int\psi(u)w(u)\,du = \int e^\top Q_*^{-1}(x-c_n u)\psi(u)w(u)g(x-c_n u)\,du$$

for $x\in\mathscr{C}_n$. Thus on the event $X\in\mathscr{C}_n$ we have

$$\begin{aligned}(\Delta_n(X)-1)^2 &= \Big(\int e^\top\big[Q^{-1}(X-c_n u) - Q_*^{-1}(X-c_n u)\big]\psi(u)w(u)g(X-c_n u)\,du\Big)^2 \\ &\le M^2\int_{[-1,1]^m}\|Q^{-1}(X-c_n u) - Q_*^{-1}(X-c_n u)\|^2\,du.\end{aligned}$$

Therefore we have

$$\begin{aligned}R_n &\le M^2\sup_{x\in\mathscr{C}} g(x)\int_{\mathscr{C}}\|Q^{-1}(x)-Q_*^{-1}(x)\|^2\,dx \\ &\le M^2\sup_{x\in\mathscr{C}} g(x)\sup_{x\in\mathscr{C}}\|Q^{-1}(x)\|^2\sup_{x\in\mathscr{C}}\|Q_*^{-1}(x)\|^2\int_{\mathscr{C}}\|Q(x)-Q_*(x)\|^2\,dx.\end{aligned}$$

By the continuity of shifts in L_2 (see Theorem 9.5 in Rudin (1974)), the square-integrability of g implies

$$\int_{\mathscr{C}} \|Q(x) - Q_*(x)\|^2 \, dx \le \sup_{u \in [-1,1]^m} \|\psi(u)\|^4 \iint (g(x - c_n u) - g(x))^2 w(u) \, du \, dx \to 0.$$

Thus we have the desired $R_n \to 0$.

3. Auxiliary results

Throughout this section let $Z, Z_1, Z_2, \dots$ be independent and identically distributed k-dimensional random vectors, and, for each x in $\mathscr{C}$, let h_{nx} be a bounded measurable function from $\mathbb{R}^k$ into $\mathbb{R}$.

Proposition 1. *Suppose that*

$$\sup_{x \in \mathscr{C}} \|h_{nx}\|_\infty = O\Big(\frac{n}{\log n}\Big), \tag{3.1}$$

$$\sup_{x \in \mathscr{C}} E[h_{nx}^2(Z)] = O\Big(\frac{n}{\log n}\Big), \tag{3.2}$$

and, for positive numbers κ_1, κ_2 and A,

$$\|h_{ny} - h_{nx}\|_\infty \le A n^{\kappa_2} \|y - x\|^{\kappa_1}, \quad x, y \in \mathscr{C}. \tag{3.3}$$

Then

$$\sup_{x \in \mathscr{C}} \Big| \frac{1}{n} \sum_{j=1}^n h_{nx}(Z_j) - E[h_{nx}(Z)] \Big| = O_p(1). \tag{3.4}$$

Proof. Let $H_n(x)$ denote the expression inside the absolute value in (3.4). We use an inequality of Hoeffding (1963): If $\xi_1, \dots, \xi_n$ are independent random variables that have mean zero and variance σ^2 and are bounded by M, then for $\eta > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{j=1}^n \xi_j\Big| \ge \eta\Big) \le 2\exp\Big(-\frac{n\eta^2}{2\sigma^2 + (2/3)M\eta}\Big).$$

Applying this inequality with $\xi_j = h_{nx}(Z_j) - E[h_{nx}(Z)]$, we obtain for $\eta > 0$:

$$P(|H_n(x)| \ge \eta) \le 2\exp\Big(-\frac{n\eta^2}{2E[h_{nx}^2(Z)] + 2\eta\|h_{nx}\|_\infty}\Big).$$

Thus there is a positive number a such that for all $\eta > 0$,

$$\sup_{x \in \mathscr{C}} P(|H_n(x)| \ge \eta) \le 2\exp\Big(-\frac{\eta^2}{1 \vee \eta} a \log n\Big).$$

Let ν be an integer greater than $\alpha + \kappa_2/\kappa_1$. We can cover $\mathscr{C}$ with $n^{\nu m}$ boxes of side lengths $n^{-\nu}$. Let C_n denote the set of centers of these smaller boxes. The above yields for $\eta > \max(1, \nu m/a)$,

$$P\Big(\max_{x\in C_n}|H_n(x)| > \eta\Big) \le \sum_{x\in C_n} P(|H_n(x)| > \eta) \le 2n^{m\nu}\exp\Big(-a\eta\log n\Big) = o(1).$$

This shows that

$$H_{n,1} = \max_{x\in C_n}|H_n(x)| = O_p(1).$$

It follows from (3.3) and $\nu > \alpha + \kappa_2/\kappa_1$ that

$$H_{n,2} = \max_{x\in C_n}\sup_{|x'-x|\le\sqrt{m}n^{-\nu}}|H_n(x') - H_n(x)| \le An^{\kappa_2}B_n^{\kappa_1}m^{\kappa_1/2}n^{-\nu\kappa_1} = o_p(1).$$

In view of the inequality $|H_n(x)| \le H_{n,1} + H_{n,2}$ for $x \in \mathscr{C}$, we have the desired result (3.4). □

In the following corollary we interpret $1/\beta$ as zero if β is infinity. We also write $\|T\|_\beta$ for the $L_\beta(P)$-norm of a random variable so that $\|T\|_\beta$ equals $(E[|T|^\beta])^{1/\beta}$ if $1 \le \beta < \infty$ and equals the essential supremum of T if $\beta = \infty$.

Corollary 1. *Suppose the function v on $\mathbb{R}^m$ is bounded, integrable and Hölder with positive exponent κ, the m-dimensional random vector X has a bounded density g, the random variable T satisfies $\|T\|_\beta < \infty$ for some $2 < \beta \le \infty$, and τg is bounded, where $\tau(X) = E(T^2|X)$. Let $c_n \to 0$ and $c_n^{-m}n^{-1+2/\beta}\log n = O(1)$. Then, for i.i.d. copies (T_j, X_j) of (T, X), we have*

$$\sup_{x\in\mathscr{C}}\Big|\frac{1}{nc_n^m}\sum_{j=1}^n\Big(T_jv\Big(\frac{X_j-x}{c_n}\Big) - E\Big[Tv\Big(\frac{X-x}{c_n}\Big)\Big]\Big)\Big| = O_p(\zeta_n^{-1/2})$$

with $\zeta_n = nc_n^m/\log n$.

Proof. Set $K = 2\|T\|_\beta$. Define

$$R_{nj}(x) = \zeta_n^{1/2}T_j\mathbf{1}[|T_j| \le Kn^{1/\beta}]\frac{1}{c_n^m}v\Big(\frac{X_j-x}{c_n}\Big),$$
$$S_{nj}(x) = \zeta_n^{1/2}T_j\mathbf{1}[|T_j| > Kn^{1/\beta}]\frac{1}{c_n^m}v\Big(\frac{X_j-x}{c_n}\Big).$$

It suffices to show that

$$\sup_{x\in\mathscr{C}}\Big|\frac{1}{n}\sum_{j=1}^n(R_{nj}(x) - E[R_{nj}(x)])\Big| = O_p(1) \tag{3.5}$$

and

$$\sup_{x\in\mathscr{C}}\Big|\frac{1}{n}\sum_{j=1}^n(S_{nj}(x) - E[S_{nj}(x)])\Big| = o_p(1). \tag{3.6}$$

Statement (3.6) is true for $\beta = \infty$ as then $S_{nj}(x) = 0$. For $\beta < \infty$ we have

$$P\Big(\max_{1\le j\le n} |T_j| > Kn^{1/\beta}\Big) \le \sum_{j=1}^n P(|T_j| > Kn^{1/\beta}) \\ \le K^{-\beta} E[|T|^\beta \mathbf{1}[|T| > Kn^{1/\beta}]] \to 0$$

and thus

$$P\Big(\sup_{x\in\mathscr{C}} \Big|\frac{1}{n}\sum_{j=1}^n S_{nj}(x)\Big| > 0\Big) \le P\Big(\max_{1\le j\le n} |T_j| > Kn^{1/\beta}\Big) \to 0.$$

In view of the inequality $E[|T|\mathbf{1}[|T| > Kn^{1/\beta}]] \le E[|T|^\beta](Kn^{1/\beta})^{1-\beta}$, we also have

$$\sup_{x\in\mathscr{C}} \Big|\frac{1}{n}\sum_{j=1}^n E[S_{nj}(x)]\Big| \le \zeta_n^{1/2} c_n^{-m} \|v\|_\infty E[|T|\mathbf{1}[|T| > Kn^{1/\beta}]] \\ = O(n^{-1/2+1/\beta} c_n^{-m/2} (\log n)^{-1/2}) = o(1).$$

This shows that (3.6) holds for $\beta < \infty$ as well.

To show (3.5) we apply Proposition 1 with $h_{nx}(T_j, X_j) = R_{nj}(x)$. We have

$$\sup_{x\in\mathscr{C}} \|h_{nx}\|_\infty \le K\|v\|_\infty n^{1/2+1/\beta} c_n^{-m/2} (\log n)^{-1/2} = O\Big(\frac{n}{\log n}\Big).$$

Furthermore,

$$\sup_{x\in\mathscr{C}} E[h_{nx}^2(T,X)] \le \frac{n}{c_n^m \log n} E\Big[\tau(X) v^2\Big(\frac{X-x}{c_n}\Big)\Big] \\ = \frac{n}{\log n}\int v^2(y)\tau(x+c_n y) g(x+c_n y)\,dy \\ \le \frac{n}{\log n}\|\tau g\|_\infty \int v^2(y)\,dy.$$

Since v is Hölder with exponent κ, we obtain, with Λ denoting the Hölder constant,

$$\|h_{ny} - h_{nx}\|_\infty \le \Big(\frac{nc_n^m}{\log n}\Big)^{1/2} Kn^{1/\beta} c_n^{-m-\kappa} \Lambda |y-x|^\kappa \le Cn^{\kappa_2}|y-x|^\kappa,$$

for some $\kappa_2 > 1 + \kappa(1+\delta-2/\beta)/m$. Thus the assumptions of Proposition 1 hold, and we obtain (3.5). □

References

[1] W. Hoeffding (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, 13–30.

[2] H. L. Koul (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models, 2nd ed.*, Lecture Notes in Statistics 166, Springer, New York.

[3] U. U. Müller, A. Schick and W. Wefelmeyer (2007). Estimating the error distribution in semiparametric regression, *Statist. Decisions* **25**, 1–18.

[4] W. Rudin (1974). *Real and Complex Analysis, 2nd ed.*, McGraw-Hill, New York.

[5] A. W. van der Vaart and J. A. Wellner (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*, Springer, New York.

Ursula U. Müller, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA

Anton Schick, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA

Wolfgang Wefelmeyer, Mathematical Institute, University of Cologne, Weyertal 86–90, 50931 Köln, Germany