JOURNAL OF MULTIVARIATE ANALYSIS 32, 213–229 (1990)

Bootstrap Approximation of Nearest Neighbor Regression Function Estimates

GERHARD DIKTA

University of Giessen, Giessen, Federal Republic of Germany

Communicated by the Editors

Let (X, Y) be a random vector in the plane and denote by $m(x) = \mathbb{E}(Y \mid X = x)$ the corresponding regression function. We show that the bootstrap approximation for the distribution of a smoothed nearest neighbor estimate of $m(x)$ is valid. Also we compare, by Monte Carlo, confidence intervals which are obtained from both the normal and the bootstrap approximation. © 1990 Academic Press, Inc.

1. INTRODUCTION AND MAIN RESULTS.

Let us assume that (X, Y) is a random vector in the plane with distribution function H. Furthermore, denote the distribution (resp. distribution function (d.f.)) of X by μ (resp. F). If $\mathbb{E}|Y| < \infty$, $m(X) := \mathbb{E}(Y|X)$, the conditional expectation of Y given X is well defined. In order to estimate the regression function, at a point x_0, we use the smoothed nearest neighbor type estimate

$$\begin{aligned} m_n(x_0) &:= (na_n)^{-1} \sum_{i=1}^{n} Y_i K[a_n^{-1}(F_n(x_0) - F_n(X_i))] \\ &= a_n^{-1} \int y K[a_n^{-1}(F_n(x_0) - F_n(x))]\, H_n(dx, dy). \end{aligned} \tag{1.1}$$

Here $(X_i, Y_i)_{i=1}^n$ is an i.i.d. sample with d.f. H, F_n is the empirical d.f. of the X sample, H_n is the bivariate empirical d.f. of the (X, Y) sample, K is a kernel function, and $(a_n)_{n \in \mathbb{N}}$ is a sequence of bandwidths tending to zero at appropriate rates. (1.1) was introduced by Yang [11] who studied

Received November 30, 1988; revised June 30, 1989.

AMS 1980 subject classifications: Primary 62E20, 62J02; Secondary 62E25, 62G05, 62G15.

Key words and phrases: nearest neighbor estimates, regression function, bootstrap approximation.

0047-259X/90 $3.00

mean square convergence of $m_n(x_0)$ to $m(x_0)$. Stute [8] showed, under appropriate conditions, that

$$(na_n)^{1/2}\,(m_n(x_0)-m(x_0)) \tag{1.2}$$

has the same normal limit distribution $\mathcal{N}(\gamma_0, \sigma_0^2)$ as

$$(na_n)^{1/2}\,a_n^{-1}\int (y-m(x_0))\,K[a_n^{-1}(F(x_0)-F(x))]\,(H_n-H)(dx, dy)+\gamma_0, \tag{1.3}$$

where

$$\gamma_0 := \frac{\sqrt{r}}{2}\,(m\circ F^{-1})''\,(F(x_0))\cdot\int_0^1 u^2K(u)\,du \qquad \text{if} \quad na_n^5\to r \tag{1.4}$$

and

$$\sigma_0^2 := \operatorname{Var}(Y\,|\,X=x)\int K^2(u)\,du,$$

and $F^{-1}(u) := \inf\{x : F(x)\geqslant u\}$ denotes the quantile function of F. Furthermore, it was pointed out that

$$m_{n0}(x_0) := m_n(x_0)\left\{a_n^{-1}\int K[a_n^{-1}(F_n(x_0)-F_n(x))]\,F_n(dx)\right\}^{-1} \tag{1.5}$$

is superior to $m_n(x_0)$ in small sample situations and that (1.5) has the same limit distribution as $m_n(x_0)$.

In order to construct a nonparametric asymptotic confidence interval for $m(x_0)$, we can use (1.2) with $m_n(x_0)$ replaced by $m_{n0}(x_0)$. In a small sample situation the limit distribution is not in any case a good approximation of the d.f. of (1.2). If we consider for example a point x_0 which is near the extreme values of the X-sample, boundary effects occur, which disappear for $n\to\infty$. These effects cannot be handled by the limit distribution. Furthermore, there is an error between (1.2) and (1.3) which also disappears for $n\to\infty$.

In this article we investigate the use of Efron's bootstrap for an approximation of the d.f. of (1.2), cf. Singh [6] for the ordinary sample mean. Since the bootstrap mimics the situation at n fixed, it may treat the small sample errors better than the limit distribution. To use the bootstrap, we first have to define the appropriate resampling scheme. Since the design is stochastic and since no assumptions on the error terms are made, the resampling has to be done on the observed data $(X_i, Y_i)_{i=1}^n$; cf. Freedman [3] for the linear correlation model.

So, let $(X_i^*, Y_i^*)_{i=1}^n$ be an i.i.d. sample with d.f. H_n and denote by H_n^* the bivariate d.f. of the (X^*, Y^*) sample, and by F_n^* the d.f. of the X^* sample. The bootstrap version of $m_n(x_0)$ is then given by

$$m_n^*(x_0) := a_n^{-1} \int yK[a_n^{-1}(F_n^*(x_0) - F_n^*(x))]\, H_n^*(dx, dy) \tag{1.6}$$

and of $m_{n0}(x_0)$ by

$$m_{n0}^*(x_0) := m_n^*(x_0) \left\{ a_n^{-1} \int K[a_n^{-1}(F_n^*(x_0) - F_n^*(x))]\, F_n^*(dx) \right\}^{-1}. \tag{1.7}$$

Denote by $\mathbb{P}^*$ the probability measure corresponding to the bootstrap sample and set

$$\bar{m}_n(x_0) := a_n^{-1} \int yK[a_n^{-1}(F(x_0) - F(x))]\, H(dx, dy).$$

We now state our main results.

THEOREM. *Assume that F is continuous, $\mathbb{E}(Y^4) < \infty$, and $m \circ F^{-1}$ is continuous on some open $U \subset (0, 1)$. Furthermore,*

> *$K: \mathbb{R} \to \mathbb{R}^+$ is twice continuously differentiable, vanishing outside the interval $[-a, a]$, $a > 0$, symmetric around zero, and strongly decreasing on $[0, a]$.* (1.8)

Furthermore, let $(a_n)_n$ be such that

$$a_n \ln(n) \to 0, \quad na_n^4/\ln(n) \to \infty, \tag{1.9}$$

$$\sum_n a_n^{-4} n^{-2} < \infty, \quad \sum_n \{\ln(n)\, n^{-1} a_n^{-1}\}^{3/2} < \infty.$$

Then with probability one, for μ a.e. x_0 with $F(x_0) \in U$,

$$\sup_{z \in \mathbb{R}} |\mathbb{P}^*[(na_n)^{1/2}(m_n^*(x_0) - m_n(x_0)) \leqslant z] - \mathbb{P}[(na_n)^{1/2}(m_n(x_0) - \bar{m}_n(x_0)) \leqslant z]| \to 0.$$

The next corollaries show that the bootstrap approximation of (1.2) holds.

COROLLARY 1. *In addition to the conditions of the theorem assume that $\int K(u)\, du = 1$, $m \circ F^{-1}$ is twice continuously differentiable on U, and $na_n^5 \to 0$.*

Then, with probability one for μ a.e. x_0 *with* $F(x_0) \in U$,

$$\sup_{z \in \mathbb{R}} |\mathbb{P}^*[(na_n)^{1/2} (m_n^*(x_0) - m_n(x_0)) \leqslant z] - \mathbb{P}[(na_n)^{1/2} (m_n(x_0) - m(x_0)) \leqslant z]| \to 0.$$

COROLLARY 2. *Under the condition of the preceding corollary, with probability one for* μ a.e. x_0 *with* $F(x_0) \in U$,

$$\sup_{z \in \mathbb{R}} |\mathbb{P}^*[(na_n)^{1/2} (m_{n0}^*(x_0) - m_{n0}(x_0)) \leqslant z] - \mathbb{P}[(na_n)^{1/2} (m_{n0}(x_0) - m(x_0)) \leqslant z]| \to 0.$$

The assumption $na_n^5 \to 0$ makes sure that the deterministic error, $(na_n)^{1/2} (\bar{m}_n(x_0) - m(x_0))$ tends to zero. If we choose an optimal sequence of bandwidths, $na_n^5 \to r > 0$, the deterministic error tends to γ_0. Since the d.f. of $(na_n)^{1/2} (m_{n0}^*(x_0) - m_{n0}(x_0))$ only approximates the d.f. of the stochastic part, $(na_n)^{1/2} (m_{n0}(x_0) - \bar{m}_n(x_0))$, we have to add an extra error term to the bootstrap, if we use optimal bandwidths; cf. Härdle and Bowman [4].

COROLLARY 3. *Let* $e_n(x_0)$ *tend to* $(m \circ F^{-1})'' (F(x_0))$ *a.s. for* μ *a.e.* x_0 *with* $F(x_0) \in U$. *Then, under the conditions of Corollary* 1 *with* $na_n^5 \to 0$ *replaced by* $na_n^5 \to r > 0$, *with probability one for* μ *a.e.* x_0 *with* $F(x_0) \in U$,

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}^* \left[(na_n)^{1/2} \left[m_{n0}^*(x_0) - m_{n0}(x_0) + a_n^2 e_n(x_0) \int u^2 K(u) \, du/2 \right] \leqslant z \right] - \mathbb{P}[(na_n)^{1/2} (m_{n0}(x_0) - m(x_0)) \leqslant z] \right| \to 0.$$

Remark. One possible candidate for $e_n(x_0)$ is again a smoothed nearest neighbor estimate

$$b_n^{-3} \int y\tilde{K}[b_n^{-1}(F_n(x_0) - F_n(x))] \, H_n(dx, dy),$$

where the kernel $\tilde{K}$ should be chosen such that $\int \tilde{K}(u) \, du = \int u\tilde{K}(u) \, du = 0$ and $\int u^2 \tilde{K}(u) \, du = 2$, and the bandwidths should tend to zero at an appropriate rate.

2. LEMMAS AND PROOFS.

In the following we denote by $\alpha_n(x) := n^{1/2}(F_n(x) - F(x))$ the empirical process, and by $\alpha_n^*(x) := n^{1/2}(F_n^*(x) - F_n(x))$ its bootstrap version.

Let $(\eta_i)_{i=1}^n$ be i.i.d. and uniformly distributed on the unit interval. Denote by $\bar{F}_n$, $\bar{\alpha}_n$ (resp. $\bar{F}_n^{-1}$) the empirical d.f., the empirical process (resp. the empirical quantile function corresponding to $(\eta_i)_{i=1}^n$. Throughout the proofs we shall often use the uniform representation

$$F_n = \bar{F}_n \circ F, \qquad \alpha_n = \bar{\alpha}_n \circ F, \qquad F_n^{-1} = F^{-1} \circ \bar{F}_n^{-1}.$$

For the bootstrap quantities we have to choose $(\eta_i)_{i=1}^n$ independently of $(X_i)_{i=1}^n$ to get

$$\mathrm{F}_n^* = \bar{F}_n \circ F_n, \qquad \alpha_n^* = \bar{\alpha}_n \circ F_n, \qquad F_i^{*-1} = F_n^{-1} \circ \bar{F}_n^{-1}.$$

When we use this representation we denote the underlying probability by $\bar{\mathbb{P}}$.

Throughout this section we assume w.l.o.g. that K has support $[-1, 1]$. For x_0 fixed, we set

$$K_n(G, x) := K[a_n^{-1}(G(x_0) - x)], \qquad K_n'(G, x) := K'[a_n^{-1}(G(x_0) - x)],$$

and

$$L_n(G, c) := \{x : |G(x_0) - G(x)| \leqslant c a_n\},$$

where G is an arbitrary distribution function and $c > 0$.

In the first two lemmas we recall some results for the oscillation behavior of α_n^*.

LEMMA 1. *Suppose that $a_n \to 0$, $na_n^2/\ln(n) \to \infty$ and $c_i > 0$ for $i = 1, 2$. Then for some constant $c > 0$,*

(i) $\mathbb{P}[\limsup_{n\to\infty} (a_n \ln(n))^{-1/2}$
$\sup_{x \in L_n(F_n, c_1) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)| < c] = 1$

(ii) $\mathbb{P}^*[(a_n \ln(n))^{-1/2}$
$\sup_{x \in L_n(F_n^*, c_1) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)| < c] \to 1$ *a.s.*

(iii) $\mathbb{P}^*[(a_n \ln(n))^{-1/2}$
$\sup_{x \in L_n(F_n^*, c_1) \cup L_n(F, c_2)} |\alpha_n^*(x_0) - \alpha_n^*(x)| < c] \to 1$ *a.s.*

Proof. Since $L_n(F_n, c_1) \subset \{x : |F(x_0) - F(x)| \leqslant c_1 a_n + 2\|F_n - F\|\}$, we have that

$$\mathbb{P}[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n, c_1) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)| > c]$$
$$\leqslant \mathbb{P}[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F, (2 + c_1 + c_2))} |\alpha_n(x_0) - \alpha_n(x)| > c]$$
$$+ \mathbb{P}(\|F_n - F\| > a_n).$$

Assertion (i) now follows from the Borel–Cantelli lemma, if we apply Lemma 2.3 in Stute [7] to the first term and the Dvoretzky–Kiefer–Wolfowitz [1] bound to the second. Using the uniform representation of the bootstrap quantities, we get that $L_n(F_n^*, c_1) \subset \{x : |F_n(x_0) - F_n(x)| \leqslant c_1 a_n + 2\|\bar{F}_n - id\|\}$ and, therefore,

$$\begin{aligned}
&\mathbb{P}^*[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n^*, c_1) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)| > c] \\
&\quad \leqslant \bar{\mathbb{P}}[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n, (2+c_1)) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)| > c] \\
&\qquad + \bar{\mathbb{P}}(\|\bar{F}_n - id\| > a_n).
\end{aligned}$$

As $(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n, (2+c_1)) \cup L_n(F, c_2)} |\alpha_n(x_0) - \alpha_n(x)|$ is deterministic with respect to $\bar{\mathbb{P}}$, (ii) follows from (i) and the DKW bound [1].

The LIL, Serfling [5, Theorem B, p. 62], yields that $\|F_n - F\|/a_n \to 0$ a.s. Therefore, we obtain a.s. that $L_n(F, c_2) \subset \{x : |F_n(x_0) - F_n(x)| < 2\|F_n - F\| + c_2 a_n\} \subset L_n(F_n, (2 + c_2))$ for n sufficiently large. The remaining part of the lemma follows from

$$\begin{aligned}
&\mathbb{P}^*[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n^*, c_1) \cup L_n(F, c_2)} |\alpha_n^*(x_0) - \alpha_n^*(x)| > c] \\
&\quad \leqslant \bar{\mathbb{P}}[(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F_n, (2+c_1+c_2))} |\bar{\alpha}_n(F_n(x_0)) - \bar{\alpha}_n(F_n(x))| > c] \\
&\qquad + \bar{\mathbb{P}}(\|\bar{F}_n - id\| > a_n),
\end{aligned}$$

Lemma 2.3 in Stute [7] and the DKW inequality [1]. ∎

LEMMA 2. *Under the conditions of the last lemma, we have, with probability one, that for each $\varepsilon > 0$ there exists $C_\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$,*

$$\mathbb{P}^*[a_n^{-1/2} \sup_{u \in U_n} |\alpha_n^*(x_0) - \alpha_n^*(F_n^{-1}(u))| > C_\varepsilon] < \varepsilon,$$

where

$$U_n := \{u : 0 < u < 1, |F(x_0) - F(F_n^{-1}(u))| < c_1 a_n\} \cup \{u : 0 < u < 1, |F(x_0) - u| < c_2 a_n\}.$$

Proof. As in Lemma 1, part (iii). ∎

Since F is continuous, $X_1, ..., X_n$ are distinct with probability one. Therefore, the resampling on the observed data $(X_i, Y_i)_{i=1}^n$ may be done in the following way. Obtain X^* according to F_n and take $J_n^*(X^*)$ as the corresponding Y^* value, where $J_n^*(x) = \sum_{i=1}^n Y_i \cdot 1_{\{X_i\}}(x)$. Then (X^*, Y^*) has the d.f. H_n and we get the representation

$$m_n^*(x_0) = a_n^{-1} \int J_n^*(x) K_n(F_n^*, F_n^*(x)) F_n^*(dx).$$

By Taylor's expansion of K at the point $(F(x_0)-F(x))/a_n$, we have that

$$\begin{aligned} m_n^*(x_0) = a_n^{-1} \int & J_n^*(x)\, K_n(F, F(x))\, F_n^*(dx) \\ & + a_n^{-2} \int J_n^*(x)[F_n^*(x_0) - F_n^*(x) - F(x_0) + F(x)] \\ & \times K_n'(F, F(x))\, F_n^*(dx) \\ & + a_n^{-3} \int J_n^*(x)[F_n^*(x_0) - F_n^*(x) - F(x_0) + F(x)]^2 \\ & \times K''(\Delta_n^*(x))/2 F_n^*(dx) \\ =: & \, I_1^* + I_2^* + I_3^*, \end{aligned}$$

and

$$\begin{aligned} m_n(x_0) = a_n^{-1} \int & y K_n(F, F(x))\, H_n(dx, dy) \\ & + n^{-1/2} a_n^{-2} \int y(\alpha_n(x_0) - \alpha_n(x))\, K_n'(F, F(x))\, H_n(dx, dy) \\ & + n^{-1} a_n^{-3} \int y(\alpha_n(x_0) - \alpha_n(x))^2\, K''(\Delta_n(x))\, H_n(dx, dy) \\ =: & \, I_1 + I_2 + I_3, \end{aligned}$$

where $\Delta_n^*(x)$ is between $(F_n^*(x_0) - F_n^*(x))/a_n$ and $(F(x_0) - F(x))/a_n$, and $\Delta_n(x)$ between $(F_n(x_0) - F_n(x))/a_n$ and $(F(x_0) - F(x))/a_n$.

LEMMA 3. *$(na_n)^{1/2} I_3 \to 0$ with probability one as $n \to \infty$.*

Proof. Since $\{x : K''(\Delta_n(x)) \neq 0\} \subset L_n(F_n, 1) \cup L_n(F, 1) =: L_n$, we have that

$$\begin{aligned} |(na_n)^{1/2} I_3| \leqslant \ln(n)(na_n^3)^{-1/2} & \left\{ (a_n \ln(n))^{-1/2} \sup_{x \in L_n} |\alpha_n(x_0) - \alpha_n(x)| \right\}^2 \\ & \times \int \|y\|\, \|K''\|/2 H_n(dx, dy). \end{aligned}$$

The assertion now follows from Lemma 1(i), $\ln(n)(na_n^3)^{-1/2} \to 0$, and the SLLN. ∎

LEMMA 4. *For almost all sample sequences and each $\varepsilon > 0$,*

$$\mathbb{P}^*[(na_n)^{1/2} |I_3^*| > \varepsilon] \to 0 \qquad \text{as} \quad n \to \infty.$$

Proof. Again we have that $\{x : K''(\Delta_n^*(x)) \neq 0\} \subset L_n(F_n^*, 1) \cup L_n(F, 1) =: L_n^*$, hence

$$(na_n)^{1/2}\, |I_3^*| \leqslant (na_n^5)^{-1/2} \int_{L_n^*} |J_n^*(x)|$$

$$\times \{(\alpha_n^*(x_0) - \alpha_n^*(x)) + (\alpha_n(x_0) - \alpha_n(x))\}^2\, \|K''\|/2F_n^*(dx)$$

$$\leqslant \ln(n)(na_n^3)^{-1/2}\, \|K''\| \int |J_n^*(x)|\, F_n^*(dx)$$

$$\times \{[(a_n \ln(n))^{-1/2} \sup_{x \in L_n^*} |\alpha_n^*(x_0) - \alpha_n^*(x)|]^2$$

$$+ [(a_n \ln(n))^{-1/2} \sup_{x \in L_n^*} |\alpha_n(x_0) - \alpha_n(x)|]^2\},$$

where we used $(a+b)^2 \leqslant 2(a^2+b^2)$ for the last inequality. By Lemma 1(i) and (ii), together with $\ln(n) \cdot (na_n^3)^{-1/2} \to 0$, the proof is complete if we show that $\mathbb{P}^*[|\int |J_n^*(x)|\, F_n^*(dx) - \mathbb{E}(|Y|)| > \varepsilon] \to 0$ a.s. But this is an easy consequence of Chebychev's inequality and the SLLN. ∎

In the following six lemmas we deal with $(na_n)^{1/2}\, (I_2^* - I_2)$.

LEMMA 5. *For almost all sample sequences and each $\varepsilon > 0$,*

$$\mathbb{P}^*\left[(na_n)^{1/2} \left| (I_2^* - I_2) - n^{-1/2} a_n^{-2} \times \int J_n^*(x)(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))\, F_n^*(dx) \right| > \varepsilon \right] \to 0.$$

Proof. As

$$I_2^* - n^{-1/2} a_n^{-2} \int J_n^*(x)(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))\, F_n^*(dx)$$

$$= n^{-1/2} a_n^{-2} \int J_n^*(x)(\alpha_n(x_0) - \alpha_n(x))\, K_n'(F, F(x))\, F_n^*(dx),$$

which is a sum of i.i.d. r.v. with ($\mathbb{P}^*-$) expectation I_2, Chebychev's inequality yields that the probability above is bounded by

$$\varepsilon^{-2} a_n^{-3} n^{-1} \mathbb{E}^*[\{J_n^*(X^*)(\alpha_n(x_0) - \alpha_n(X^*))\, K_n'(F, F(X^*))\}^2]$$

$$\leqslant \varepsilon^{-2} n^{-1} a_n^{-2} \ln(n)\, \{(a_n \ln(n))^{-1/2} \sup_{x \in L_n(F, 1)} |\alpha_n(x_0) - \alpha_n(x)|\}^2$$

$$\times \|K'\|^2 \cdot n^{-1} \sum_{i=1}^{n} Y_i^2.$$

The last term tends to 0 by the SLLN, Lemma 1(i) and (1.9). ∎

The last lemma allows us to replace $(na_n)^{1/2}(I_2^* - I_2)$ by

$$a_n^{-3/2} \int J_n^*(x)(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))\, F_n^*(dx).$$

In the following lemma we prove that the measure F_n^* can be replaced by F_n.

LEMMA 6. *For almost all sample sequences and each* $\varepsilon > 0$,

$$\mathbb{P}^*\left[a_n^{-3/2}\left|\int J_n^*(x)(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))(F_n^* - F_n)(dx)\right| > \varepsilon\right] \to 0.$$

Proof. Define

$$k_n^*(x, z) := J_n^*(x)\, K_n'(F, F(x))\{1_{(-\infty, x_0]}(z) - 1_{(-\infty, x]}(z)\},$$

to get

$$J_n^*(x)(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x)) = n^{1/2} \int k_n^*(x, z)(F_n^* - F_n)(dz)$$

and therefore, by Markov's inequality, as a bound for the probability above

$$\varepsilon^{-2} a_n^{-3} n \mathbb{E}^*\left[\left\{\int\int k_n^*(x, z)(F_n^* - F_n)(dz)(F_n^* - F_n)(dx)\right\}^2\right] =: \varepsilon^{-2} a_n^{-3} n c_n.$$

Since $(na_n^3)^{-1} \to 0$ it remains to prove that $n^2 c_n$ is bounded a.s. As in the proof of Lemma B on page 223 in Serfling [5] with the special V-statistic $\int k_n^*(x, z)(F_n^* - F_n)(dz)(F_n^* - F_n)(dx)$, we can bound c_n, the second ($\mathbb{P}^* -$) moment of the V-statistic, by $n^{-2} c\{n^{-1} \sum_{i=1}^n Y_i^2 + [n^{-1} \sum_{i=1}^n |Y_i|]^2\}$, where c is a constant, depending only on $\|K'\|$. The SLLN then proves the lemma. ∎

The next lemma justifies the replacement of $J_n^*(x)$ by $m(x)$.

LEMMA 7. *With probability one and* μ *a.e.* x_0, *for each* $\varepsilon > 0$,

$$\mathbb{P}^*\left[a_n^{-3/2}\left|\int (J_n^*(x) - m(x))(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))\, F_n(dx)\right| > \varepsilon\right] \to 0.$$

Proof. By Markov's inequality the above probability is less than or equal to

$$\varepsilon^{-2} a_n^{-3} n^{-2} \sum_{i,j=1}^{n} B_i B_j K_i K_j \mathbb{E}^*(A_i^* A_j^*) =: \varepsilon^{-2} L_n,$$

where we have set $A_i^* := \alpha_n^*(x_0) - \alpha_n^*(X_i)$, $B_i := Y_i - m(X_i)$, and $K_i := K_n'(F, F(X_i))$. To show that $\limsup L_n = 0$ a.s. for μ a.e. x_0, we apply Borel–Cantelli. First observe that

$$\mathbb{E}(B_i B_j B_k B_l \cdot K_i K_j K_k K_l \mathbb{E}^*(A_i^* A_j^*)\, \mathbb{E}^*(A_k^* A_l^*)) = 0$$

if there is one index in the sequence of indices i, j, k, l, which is contained only once in the sequence. With $|\mathbb{E}^*(A_i^* A_j^*)| \leq 8$ we therefore obtain, again by Markov's inequality, that for $\delta > 0$

$$\mathbb{P}(L_n > \delta) \leq c a_n^{-6} n^{-4} \{ n \|K'\|^4\, \mathbb{E}(B_1^4) + n(n-1)\, \mathbb{E}(B_1^2 K_1^2)^2 \},$$

where $c > 0$ is an appropriate constant depending only on δ. By (1.9) it remains to show that $\limsup a_n^{-2} \mathbb{E}(B_1^2 K_1^2)^2 < \infty$ for μ a.e. x_0. Since $F(X)$ is uniformly distributed on $[0, 1]$, we get

$$\begin{aligned} a_n^{-1} \mathbb{E}(B_1^2 K_1^2) &= a_n^{-1} \mathbb{E}[\mathbb{E}(\{Y - m(X)\}^2 \,|F(X))\, K_n'(F, F(X))^2] \\ &= a_n^{-1} \int_0^1 M(u)\, K'(F, u)^2\, du \\ &\leq a_n^{-1} \|K'\|^2 \int_{(F(x_0) - a_n,\, F(x_0) + a_n)} M(u)\, du, \end{aligned}$$

where $M(u) := \mathbb{E}(\{Y - m(X)\}^2 \,|\, F(X) = u)$. Theorem 10.49 in Wheeden and Zygmund [10] yields that the last term tends to a finite constant, depending on $F(x_0)$, for Lebesgue almost all $F(x_0)$. Therefore, $\limsup a_n^{-1} \mathbb{E}(B_1^2 K_1^2) < \infty$ for μ a.e. x_0. ∎

Next, we replace $m(x)$ by $m(x_0)$.

LEMMA 8. *With probability one and μ a.e. x_0 with $F(x_0) \in U$, for each $\varepsilon > 0$*

$$\mathbb{P}^* \left[a_n^{-3/2} \left| \int (m(x) - m(x_0))(\alpha_n^*(x_0) - \alpha_n^*(x))\, K_n'(F, F(x))\, F_n(dx) \right| > \varepsilon \right] \to 0.$$

Proof. Transformation of the above integral gives

$$\mathbb{P}^* \left[a_n^{-1/2} \sup_{u \in U_n} |\alpha_n^*(x_0) - \alpha_n^*(F_n^{-1}(u))| \right.$$
$$\left. \times\, a_n^{-1} \int_0^1 |m(F_n^{-1}(u)) - m(x_0)|\, |K_n'(F, F(F_n^{-1}(u)))|\, du > \varepsilon \right]$$

as a bound for the probability in question, where

$$U_n := \{u \in (0,1) : |F(x_0) - F(F_n^{-1}(u))| < a_n\}.$$

By the mean value theorem we have that

$$a_n^{-1} \int_0^1 |m(F_n^{-1}(u)) - m(x_0)|\ |K_n'(F, F(F_n^{-1}(u))) - K_n'(F, u)|\ du$$

$$\leqslant a_n^{-2}\ \|K''\|\ \|F(F_n^{-1}) - id\| \int_0^1 |m(F_n^{-1}(u)) - m(x_0)|\ du.$$

Since $\|F(F_n^{-1}) - id\| \leqslant \|F - F_n\| + n^{-1}$, the LIL, cf. Serfling [5, Theorem B, p. 62], implies that $a_n^{-2}\|F(F_n^{-1}) - id\| \to 0$ a.s. Furthermore,

$$\int_0^1 |m(F_n^{-1}(u)) - m(x_0)|\ du \to \mathbb{E}(|m(X) - m(x_0)|) \qquad \text{a.s.}$$

by the SLLN. According to Lemma 2, it remains to show that

$$a_n^{-1} \int_0^1 |m(F_n^{-1}(u)) - m(x_0)|\ |K_n'(F, u)|\ du \to 0 \qquad \text{a.s.} \qquad \text{for} \quad \mu \text{ a.e. } x_0.$$

Use the uniform representation of F_n^{-1} to show that the integral above equals

$$\int_{-1}^1 |m(F^{-1}(\bar{F}_n^{-1}(F(x_0) - ua_n))) - m(F^{-1}(F(x_0)))|$$

$$\times |K_n'(u)|\ du \qquad \text{for} \quad \mu \text{ a.e. } x_0.$$

Observing that $\sup_{u:|u| \leqslant 1} |\bar{F}_n^{-1}(F(x_0) - ua_n) - F(x_0)| \leqslant \|\bar{F}_n^{-1} - id\| + a_n \to 0$ a.s., the assertion follows from the continuity of $m \circ F^{-1}$ at $F(x_0) \in U$. ∎

Considering the proof of Lemma 8 again, we find

LEMMA 9. *With probability one and μ a.e. x_0, for each $\varepsilon > 0$,*

$$\mathbb{P}^*\Bigg[a_n^{-3/2} \Bigg|\int_0^1 m(x_0)(\alpha_n^*(x_0) - \alpha_n^*(F_n^{-1}(u)))$$

$$\times (K_n'(F, F(F_n^{-1}(u))) - K'_n(F, u))\ du\Bigg| > \varepsilon\Bigg] \to 0.$$

LEMMA 10. *For a.e. sample sequence and μ a.e. x_0 with $F(x_0) \in U$ we have, for each $\varepsilon > 0$,*

$$\mathbb{P}^*\left[\left|a_n^{-3/2}\int_0^1 m(x_0)(\alpha_n^*(x_0)-\alpha_n^*(F_n^{-1}(u)))\right.\right.$$
$$\left.\left.\times K_n'(F,u)\,du + a_n^{-1/2}\int m(x_0)\,K_n(F,F(x))\,\alpha_n^*(dx)\right| > \varepsilon\right] \to 0.$$

Proof. For $0 < F(x_0) < 1$ choose n sufficiently large to get $\int_0^1 K_n'(F,u)\,du = 0$, $K_n(F,1) = 0$, and $K_n(F,0) = 0$. Integration by parts then implies the equality of the first integral to

$$-a_n^{-1/2} m(x_0) \int_0^1 K_n(F,u)\,d\alpha_n^*(F_n^{-1}(u)).$$

A straightforward examination of this term shows the equality to

$$-a_n^{-1/2}\,m(x_0)\int_0^1 K_n(F,F_n(x-))\,\alpha_n^*(dx),$$

where $K_n(F, F_n(x-))$ denotes the left-hand limit of $K_n(F, F_n(x))$. By Chebychev's inequality we have

$$m(x_0)^2\,\varepsilon^{-2}a_n^{-1}n^{-1}\sum_{i=1}^n \mathrm{Var}^*[K_n(F,F_n(X_i^*-)) - K_n(F,F(X_i^*))]$$

as a bound for the probability above. From the mean value theorem, this can be bounded by

$$m(x_0)^2\,\varepsilon^{-2}a_n^{-1}\,\|K'\|^2\,\{a_n^{-1}\sup_{x\in\mathbb{R}}|F(x)-F_n(x-)|\}^2.$$

Since F is continuous, $\sup_{x\in\mathbb{R}}|F(x)-F_n(x-)| = \|F_n - F\|$, we get, by LIL, $\limsup a_n^{-3/2}\sup_{x\in\mathbb{R}}|F(x)-F_n(x)| = 0$ a.s., whence the assertion. ∎

We are now in a position to give the

Proof of the theorem. By the theorem in Stute [8] and since $\mathcal{N}(0,\sigma_0^2)$ is continuous, we have to prove that $(na_n)^{1/2}(m_n^*(x_0) - m_n(x_0))$ tends to $\mathcal{N}(0,\sigma_0^2)$ in distribution for $\mathbb{P}$ almost all sample sequences and μ a.e. x_0 with $F(x_0) \in U$. Upon using the lemmas, it remains to show that for $\mathbb{P}$ almost all sample sequences

$$n^{1/2}a_n^{-1/2}\int (J_n^*(x) - m(x_0))\,K_n(F,F(x))(F_n^* - F_n)\,dx \qquad (2.1)$$

tend to $\mathcal{N}(0, \sigma_0^2)$ in distribution for μ a.e. x_0 with $F(x_0) \in U$. Define, for n fixed,

$$Z_{n,i}^* := a_n^{-1/2}\{(J_n^*(X_i^*) - m(x_0))\, K_n(F, F(X_i^*)) \\ - \mathbb{E}^*((J_n^*(X_1^*) - m(x_0))\, K_n(F, F(X_1^*)))\}$$

and observe that (2.1) is equal to $n^{-1/2} \sum_{i=1}^n Z_{n,i}^*$, where $(Z_{n,i}^*)_{i=1,\dots,n}$ are i.i.d. with zero mean. For the ($\mathbb{P}^*-$) variance of this standardized sum we get

$$a_n^{-1} \left\{ n^{-1} \sum_{i=1}^n (Y_i - m(x_0))^2 K_n^2(F, F(X_i)) \right. \\ \left. - \left[n^{-1} \sum_{i=1}^n (Y_i - m(x_0))\, K_n(F, F(X_i)) \right]^2 \right\} =: M_{n,1} - M_{n,2}.$$

Theorem 3, together with the remark on page 893 in Stute [9] implies $M_{n,1} \to \sigma_0^2$ $\mathbb{P}$ a.s. for μ a.e. x_0, where we have to observe that $\mathbb{E}((Y - m(x_0)^2 \mid F(X) = F(x_0))$ is equal to $\mathbb{E}((Y - m(x_0)^2 \mid X = x_0)$ μ a.s. By the same argument we get $M_{n,2} \to 0$ $\mathbb{P}$ a.s. for μ a.e. x_0, therefore,

$$\mathrm{Var}^* \left[n^{-1/2} \sum_{i=1}^n Z_{n,i}^* \right] \to \sigma_0^2, \qquad \mathbb{P} \text{ a.s. for } \mu \text{ a.e. } x_0. \tag{2.2}$$

It remains to verify Lindeberg's condition for the array $((Z_{n,i}^*)_{i=1}^n)_{n \in \mathbb{N}}$. This amounts to

$$\mathbb{E}^*(Z_n^{*2} \cdot 1_{\{Z_n^{*2} > \varepsilon n\}}) \to 0 \tag{2.3}$$

for each $\varepsilon > 0$, where $Z_n^* := Z_{n,1}^*$. As in the proof of (2.2),

$$n^{-1} \sum_{i=1}^n (Y_i - m(x_0))\, K_n(F, F(X_i))$$

tends to zero $\mathbb{P}$ a.s. for μ a.e. x_0. So, since $na_n \to \infty$, it suffices to show that for μ a.e. x_0 with $F(x_0) \in U$,

$$\limsup_{n \to \infty} \sum_{i=1}^n A_{n,i}^2 1_{\{Y_i^2 > \varepsilon n a_n\}} < \infty, \qquad \mathbb{P} \text{ a.s.},$$

where

$$A_{n,i} := (Y_i - m(x_0))\, K_n(F, F(X_i)) - n^{-1} \sum_{j=1}^n (Y_j - m(x_0))\, K_n(F, F(X_j)).$$

Observing $na_n^4 \to \infty$, this follows from $\mathbb{E}(|Y|^{8/3}) < \infty$, since then

$$\limsup_{i \to \infty} |Y_i|^{8/3}/i = 0, \qquad \mathbb{P} \text{ a.s. } \blacksquare$$

Proof of Corollary 1. In the preceding proof we saw that $(na_n)^{1/2}(m_n^*(x_0) - m_n(x_0))$ tends to $\mathcal{N}(0, \sigma_0^2)$ in distribution for $\mathbb{P}$ almost all sequences and μ a.e. x_0 with $F(x_0) \in U$. According to the continuity of $\mathcal{N}(0, \sigma_0^2)$ it remains to show that $(na_n)^{1/2}(m_n(x_0) - m(x_0))$ tends to $\mathcal{N}(0, \sigma_0^2)$ in distribution for μ a.e. x_0 with $F(x_0) \in U$. But this is the assertion of the corollary in Stute [8]. $\blacksquare$

Proof of Corollary 2. As mentioned in the introduction in Stute [8], $(na_n)^{1/2}(m_{n0}(x_0) - m(x_0))$ tends to $\mathcal{N}(0, \sigma_0^2)$ in distribution for μ a.e. x_0 with $F(x_0) \in U$. Defining

$$f_n(x_0) := a_n^{-1} \int K_n(F_n, F_n(x))\, F_n(dx),$$

$$f_n^*(x_0) := a_n^{-1} \int K_n(F_n^*, F_n^*(x))\, F_n^*(dx),$$

we have that $m_{n0}(x_0) = m_n(x_0)/f_n(x_0)$ and $m_{n0}^*(x_0) = m_n^*(x_0)/f_n^*(x_0)$. Since

$$\begin{aligned} &(na_n)^{1/2}(m_{n0}^*(x_0) - m_{n0}(x_0)) \\ &\quad = (na_n)^{1/2}\{(m_n^*(x_0) - m_n(x_0)) f_n(x_0) - m_n(x_0)(f_n^*(x_0) - f_n(x_0))\} \\ &\quad \times \{f_n^*(x_0) \cdot f_n(x_0)\}^{-1}, \end{aligned}$$

the desired result follows by the preceding theorem if we show that for $\mathbb{P}$ a.e. sample sequence and μ a.e. x_0 with $F(x_0) \in U$,

$$m_n(x_0) \to m(x_0), \qquad f_n(x_0) \to 1, \tag{2.4}$$

and

$$\mathbb{P}^*[(na_n)^{1/2} |f_n^*(x_0) - f_n(x_0)| > \varepsilon] \to 0, \qquad \text{for each} \quad \varepsilon > 0. \tag{2.5}$$

But (2.4) is an immediate consequence of Stute [9, Theorem 3], and (2.5) is obtained by checking the preceding lemmas with Y_i replaced by 1. $\blacksquare$

Proof of Corollary 3. Since $(na_n^5)^{1/2} e_n(x_0) \int u^2 K(u)\, du/2 \to \gamma_0$ a.s. for μ a.e. x_0 with $F(x_0) \in U$, the assertion can be derived as in the proof of Corollary 2 together with Remark 1 in Stute [8]. $\blacksquare$

3. Some Simulation Results

In the following we use Corollary 2 to construct 90% confidence intervals for $m(x_0)$. Having observed the data $(x_1, y_1), \ldots, (x_n, y_n)$, we calculate the 5% and 95% quantiles of the distribution function of $(na_n)^{1/2}(m^*_{n0}(x_0) - m_{n0}(x_0))$, and use them as an approximation of the 5% (resp. 95%) quantile of the distribution function of $(na_n)^{1/2}(m_{n0}(x_0) - m(x_0))$. Since the exact calculation of the bootstrap quantiles is too comprehensive, we use a Monte-Carlo evaluation; cf. Efron [2, Chap. 5]. For this we draw a bootstrap sample from H_n of size n and calculate $m^*_{n0}(x_0)$. This is done 1000 times, and we take $(na_n)^{1/2}((m^*_{n0}(x_0))_{(50)} - m_{n0}(x_0))$ as an estimate for the 5% bootstrap quantile (resp. $(na_n)^{1/2}((m^*_{n0}(x_0)_{(950)} - m_{n0}(x_0))$ for the 95% bootstrap quantile), where $((m^*_{n0}(x_0))_{(k)})_{k=1}^{1000}$ denote the ordered bootstrap values. After this, we compute the bootstrap confidence interval

$$[2m_{n0}(x_0) - (m^*_{n0}(x_0))_{(950)}, 2m_{n0}(x_0) - (m^*_{n0}(x_0))_{(50)}]. \tag{3.1}$$

In our simulation study we examined two models:

$$\text{Model 1}: X \sim U(-1, 1), \varepsilon \sim U(-0.5, 0.5), X \text{ independent of } \varepsilon,$$

$$Y = 5X^2 + 7X + \varepsilon; \tag{3.2}$$

$$\text{Model 2}: X \text{ and } \varepsilon \text{ as in (3.2), but}$$

$$Y = 5X^2 + 7X + (X+1)\varepsilon; \tag{3.3}$$

where $U(a, b)$ denotes the uniform distribution on (a, b). For sample size $n = 50$ we constructed, for each point $x_0 \in \{-0.7, -0.4, -0.2, 0, 0.2, 0.4, 0.7\}$, 100 intervals, using the bootstrap approximation, as described above, and 100 intervals by normal approximation, using the corollary in Stute [8] with $m_{n0}(x_0)$ in place of $m_n(x_0)$. Put

$$\hat{\sigma}_0^2 := \left\{\left\{\int y^2 K_n(F_n, F_n(x))\, H_n(dx, dy)\right\}\left\{\int K_n(F_n, F_n(x))\, F_n(dx)\right\}^{-1} - m_{n0}^2(x_0)\right\} \int K^2(u)\, du \tag{3.4}$$

as an estimate for σ_0^2. For K and a_n we used

$$K(x) := 3/4(1-x)^2\, 1_{[-1,1]}(x)$$

and

$$a_n := (n \ln(n))^{-1/5} \approx 0.35. \tag{3.5}$$

Table I shows the result.

TABLE I

	Eg. (3.2)—Model 1				Eg. (3.3)—Model 2			
	C.P.(%)		W.		C.P.(%)		W.	
x_0	N.A.	B.A.	N.A.	B.A.	N.A.	B.A.	N.A.	B.A.
−0.7	14	57	0.42	0.52	0	31	0.38	0.51
−0.4	33	58	0.91	0.9	35	65	0.84	0.85
−0.2	82	85	1.34	1.28	85	87	1.34	1.24
0	97	91	1.8	1.7	97	93	1.87	1.78
0.2	99	86	2.26	2	97	90	2.32	2.05
0.4	99	90	2.62	1.99	99	90	2.66	2.05
0.7	32	32	2.48	2.06	48	50	2.53	2.07

Note. C.P.—coverage probability of the 100 intervals; W.—average interval width of the 100 intervals; N.A.—normal approximation; B.A.—bootstrap approximation.

In both models the regression function equals

$$m(x_0) = 5x_0^2 + 7x_0, \tag{3.6}$$

while

$$\mathrm{Var}(Y \mid X = x_0) = 1/12 \text{ in model 1} \tag{3.7}$$

and

$$\mathrm{Var}(Y \mid X = x_0) = (x_0 + 1)^2/12 \text{ in model 2.} \tag{3.8}$$

The bad C.P. at the points -0.7 and 0.7 is due to boundary effects. Nevertheless, the point $x_0 = -0.7$ in Model 2 is quite interesting, since the regression function attains its minimum there, while the variance is very small. Both facts together result in skew and small intervals when we use normal approximation.

References

[1] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[2] Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. (*Reg. Conf. Ser. Appl. Math.*) *SIAM* **38**.

[3] Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

[4] HÄRDLE, W., AND BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bounds. *J. Amer. Statist. Assoc.* **83** 102–110.

[5] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[6] SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

[7] STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

[8] STUTE, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.* **12** 917–926.

[9] STUTE, W. (1986). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.* **14** 891–901.

[10] WHEEDEN, R. L., AND ZYGMUND, A. (1977). *Measure and Integral.* Dekker, New York.

[11] YANG, S. S. (1981). Linear functions of concomitants of order statistics with applications to nonparametric estimation of a regression function. *J. Amer. Statist. Assoc.* **76** 658–662.