The Annals of Statistics
1986, Vol. 14, No. 2, 638–647

CONDITIONAL EMPIRICAL PROCESSES

BY WINFRIED STUTE

University of Giessen

We prove a Donsker-type invariance principle for a nearest-neighbor-type conditional empirical process. As an application we show asymptotic normality of conditional quantiles and derive large-sample distribution-free tests and confidence bands for a conditional distribution function.

1. Introduction and main results. Let (X, Y) be a random vector in $\mathbb{R}^{1+d}$ with distribution function H. For real Y (i.e., $d = 1$) with $\mathbb{E}(|Y|) < \infty$ write $\mathbb{E}(Y|X) = m \circ X$, with $m(x) = \mathbb{E}(Y|X = x)$ denoting the regression function of Y at $X = x$. Assume that $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of independent random vectors with the same distribution as (X, Y). Much work has been devoted to the problem of (nonparametric) estimation of m when only little information on H is available. See Collomb (1981) for a survey.

For a general d, replacing Y by the indicator function $1_{\{Y \leq \mathbf{y}\}}$, $\mathbf{y} \in \mathbb{R}^d$, we might apply the existing results for statistical inference about the conditional distribution function

$$m(\mathbf{y}|x) = \mathbb{P}(Y \leq \mathbf{y}|X = x), \qquad (x, \mathbf{y}) \in \mathbb{R}^{1+d}$$

at a fixed point $\mathbf{y} \in \mathbb{R}^d$. As in the case of unconditional distribution functions, such a result is insufficient for most purposes. For example, when dealing with smooth functionals of $m(\cdot|x)$, it is necessary to handle estimates $m_n(\cdot|x) = m_n(\cdot|x; X_1, Y_1, \ldots, X_n, Y_n)$ of $m(\cdot|x)$ as a function rather than its value at a single point. In other words, it is desirable to study the distributional character of the process $\{m_n(\mathbf{y}|x): \mathbf{y} \in \mathbb{R}^d\}$.

In Stute (1984b) we introduced a nearest-neighbor-type estimate of $m(x)$, which turned out to be asymptotically normal under minimal assumptions on H. To be explicit, let $d = 1$ and write, for $n \geq 1$,

$$F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{(-\infty, x]}(X_i), \qquad x \in \mathbb{R},$$

the empirical distribution function of $X_1, \ldots, X_n$. Let K be a smooth probability kernel with bounded support and put, for some bandwidth $a_n > 0$,

$$m_n(x_0) = (na_n)^{-1} \sum_{i=1}^{n} Y_i K\left(\frac{F_n(x_0) - F_n(X_i)}{a_n}\right).$$

Under some mild growth conditions on a_n $(\to 0)$ it was shown that

$$(na_n)^{1/2}[m_n(x_0) - m(x_0)] \to N(0, \sigma^2)$$

Received July 1984; revised March 1985.

AMS 1980 *subject classifications.* Primary 60F17, 62J02; secondary 62G05, 62G10, 62G15.

Key words and phrases. Conditional empirical distribution function, invariance principle, conditional quantiles.

in distribution, where

$$\sigma^2 = \operatorname{var}(Y|X = x_0) \int K^2(u)\, du.$$

As indicated above, replacing Y_i by $1_{(-\infty, \mathbf{y}]} \circ Y_i$, we obtain the process

$$m_n(\mathbf{y}|x_0) = (na_n)^{-1} \sum_{i=1}^{n} 1_{(-\infty, \mathbf{y}]} \circ Y_i \cdot K\left(\frac{F_n(x_0) - F_n(X_i)}{a_n} \right), \qquad \mathbf{y} \in \mathbb{R}^d,$$

as an estimate of $m(\cdot|x_0)$.

The main result of this paper states that when viewed as a random element in a suitable (topological) space of functions, $(na_n)^{1/2}[m_n(\cdot|x_0) - m(\cdot|x_0)] \to B_0$ in distribution, where B_0 is a certain Gaussian process (depending on x_0). In other words, we prove a Donsker-type invariance principle for the conditional process $m_n(\mathbf{y}|x_0)$, $\mathbf{y} \in \mathbb{R}^d$.

Observe that, since K is a probability kernel, $m_n(\mathbf{y}|x_0)$ is nonnegative and "nondecreasing" in $\mathbf{y}$. It is not a proper distribution function, however, since in general the weights $K([F_n(x_0) - F_n(X_i)]/a_n)/na_n$, $1 \le i \le n$, do not sum up to one. In the next section, we shall propose a modification of m_n, which turns out to be a proper distribution function with the same asymptotic behavior as m_n.

In the following write $Y = (Y^1, \dots, Y^d)$, and denote with G^j, $1 \le j \le d$, the marginal distribution function of Y^j. Define G: $\mathbb{R}^d \to [0,1]^d$ by $G(y_1, \dots, y_d) := (G^1(y_1), \dots, G^d(y_d))$, and let F be the distribution function of X. We then have

$$H(x, y_1, \dots, y_d) = C\big(F(x), G^1(y_1), \dots, G^d(y_d)\big),$$

where C is the copula function of H, a distribution function on $[0,1]^{1+d}$ with uniform marginals. Similarly, for the empirical distribution function H_n of $(X_1, Y_1), \dots, (X_n, Y_n)$ we may write

$$H_n(x, y_1, \dots, y_d) = C_n\big(F(x), G^1(y_1), \dots, G^d(y_d)\big),$$

where C_n is the empirical distribution function of an i.i.d. sequence with distribution function C. The possibility of obtaining H and H_n from the "uniform" processes C and C_n by means of the transformation $(F, G^1, \dots, G^d) \in [0,1]^{1+d}$ is important for deriving statements about multivariate empirical processes on $\mathbb{R}^{1+d}$ from corresponding processes on $[0,1]^{1+d}$ with uniform marginals. Moreover, since the weights $K([F_n(x_0) - F_n(X_i)]/a_n)$ have the nice property of depending only on the order statistics and ranks of $X_1, \dots, X_n$, we may write

$$\begin{aligned} m_n(\mathbf{y}|x_0) &= a_n^{-1} \int 1_{(-\infty, \mathbf{y}]}(\mathbf{u}) K\left(\frac{F_n(x_0) - F_n(x)}{a_n} \right) H_n(dx, d\mathbf{u}) \\ &= a_n^{-1} \int 1_{[\mathbf{0}, G(\mathbf{y})]}(\mathbf{u}) K\left(\frac{\bar{F}_n(F(x_0)) - \bar{F}_n(x)}{a_n} \right) C_n(dx, d\mathbf{u}) \\ &\equiv \tilde{m}_n(G(\mathbf{y})|F(x_0)), \end{aligned}$$

where $\bar{F}_n$ is the first marginal distribution of C_n, an empirical distribution pertaining to an i.i.d. sample with uniform distribution. Consequently, in order to derive distributional results for m_n, we may and do assume that H has uniform marginals.

Throughout this paper assume that K is a twice continuously differentiable probability kernel vanishing outside some finite interval. $(a_n)_n$ will be a sequence of bandwidths converging to zero at appropriate rates.

While K and a_n are at the statistician's disposal, the invariance principle may be proved only under an additional smoothness assumption on the unknown H (resp. m). Recall that for m_n we may assume w.l.o.g. that F is the uniform distribution on $[0,1]$.

ASSUMPTION (A). Assume that

$$\sup_{\|\mathbf{t}-\mathbf{s}\|\le\delta} |m(\mathbf{t}|x) - m(\mathbf{s}|x)| = o\big((\ln\delta^{-1})^{-1}\big) \quad \text{as} \quad \delta\to 0$$

uniformly in a neighborhood of x_0.

Clearly (A) is satisfied whenever m is Hölder continuous of some positive order. No existence of densities is required. (A) also guarantees that m is equicontinuous in a neighborhood of x_0. This is quite natural in view of the fact that the standardized process m_n is expected to have a limit process with continuous sample paths. Now, for $\mathbf{y}\in[0,1]^d$, put

$$\overline{m}_n(\mathbf{y}|x_0) = a_n^{-1}\int 1_{[\mathbf{0},\mathbf{y}]}(\mathbf{u})K\left(\frac{x_0-x}{a_n}\right)H(dx,d\mathbf{u}).$$

Recall $F = U[0,1]$, the uniform distribution on $[0,1]$, and observe that, by definition of $m(\cdot|x)$,

$$\overline{m}_n(\mathbf{y}|x_0) = a_n^{-1}\int_0^1 m(\mathbf{y}|x)K\left(\frac{x_0-x}{a_n}\right)dx,$$

a smoothed version of $m(\mathbf{y}|x_0)$.

To state our first main result, we denote with $D[0,1]^d$ the space of all "right-continuous" functions on $[0,1]^d$ with "left-hand" limits; cf. Billingsley (1968) for $d=1$ and Neuhaus (1971) for a general d. Endow $D[0,1]^d$ with the Skorokhod topology, and let $\mathscr{B}(D)$ be the generated Borel σ field. Clearly, m_n is a random element in $(D,\mathscr{B}(D))$, so its distribution is well-defined.

THEOREM 1. *Assume that H has uniform marginals, and let $a_n\to 0$ be such that $na_n^3\to\infty$. Under (A) we then have for Lebesgue-almost all $0<x_0<1$*

$$(na_n)^{1/2}[m_n(\cdot|x_0) - \overline{m}_n(\cdot|x_0)] \to B_0 \equiv B_0(x_0) \quad \text{in distribution}.$$

Here B_0 is a centered Gaussian process on $[0,1]^d$ with continuous sample paths vanishing at the lower boundary of $[0,1]^d$ and covariance

$$\operatorname{cov}(B_0(\mathbf{y}_1),B_0(\mathbf{y}_2)) = [m(\mathbf{y}_1\wedge\mathbf{y}_2|x_0) - m(\mathbf{y}_1|x_0)m(\mathbf{y}_2|x_0)]\int K^2(u)\,du.$$

In other words, B_0 is a scaled tied-down Brownian sheet with intensity measure $m(\cdot|x_0)$. When X is independent of Y, $m(\cdot|x_0) = Q_Y$, the distribution of Y for all x_0. Hence up to a scaling factor, B_0 is equal to the limit of the unconditional empirical process pertaining to the Y sequence, as should be expected. Observe, however, that the standardizing factor is $(na_n)^{1/2}$ with $a_n \to 0$, indicating a lower rate of convergence. This is the price one has to pay when making inference about conditional (local) quantities.

It is not hard to prove that the standardized processes $m_n(\cdot|x)$ converge jointly in distribution to $B_0(x)$ even at finitely many points $x = x_1, \ldots, x_k$, with $B_0(x_1), \ldots, B_0(x_k)$ being independent.

The corresponding invariance principle for $(na_n)^{1/2}[m_n(\cdot|x_0) - m(\cdot|x_0)]$ may be obtained under an additional smoothness condition on $m(\mathbf{y}|x)$ as a function of x. This is necessary in order to guarantee that $\overline{m}_n - m \to 0$ at a satisfactory rate.

ASSUMPTION (B). For each $\mathbf{y}$ $m(\mathbf{y}|\cdot)$ is twice continuously differentiable in a neighborhood U of x_0, such that

$$\sup_{x\in U} \sup_{\mathbf{y}} |m''(\mathbf{y}|x)| < \infty.$$

COROLLARY 2. *Under the conditions of the theorem, assume that* (B) *holds, and let* K *be such that* $\int uK(u)\,du = 0$. *Whenever* $na_n^5 \to 0$ *we have for Lebesgue-almost all* $0 < x_0 < 1$

$$(na_n)^{1/2}[m_n(\cdot|x_0) - m(\cdot|x_0)] \to B_0 \quad \textit{in distribution}.$$

PROOF. According to the theorem it remains to show that $(na_n)^{1/2}[\overline{m}_n(\mathbf{y}|x_0) - m(\mathbf{y}|x_0)] \to 0$ uniformly in $\mathbf{y}$. Because of $na_n^5 \to 0$, it suffices to prove $\overline{m}_n - m = O(a_n^2)$. This follows, however, in much the same way as the corollary in Stute (1984b). □

With the same method of proof, one may also treat the optimal choice of a bandwidth, namely $na_n^5 \to c > 0$. For a general c, the limit process is equal to the noncentered Gaussian process

$$B_0^c\colon \mathbf{y} \to B_0(\mathbf{y}) + \frac{\sqrt{c}\, m''(\mathbf{y}|x_0)}{2} \int u^2K(u)\,du.$$

Clearly, for B_0^c, $c > 0$, to be continuous, we also need continuity of $m''(\cdot|x_0)$. As for the usual empirical process, the invariance principle for the conditional empirical process may be used to test the hypothesis $m(\cdot|x_0) = m_0(\cdot|x_0)$ and to determine confidence bands for $m(\cdot|x_0)$. For example, when $d \equiv 1$ and G is continuous, we have (when $c = 0$)

$$\begin{aligned}
&(na_n)^{1/2} \sup_{y\in\mathbb{R}} |m_n(y|x_0) - m(y|x_0)| \\
&\quad = (na_n)^{1/2} \sup_{0\le u\le 1} |\tilde{m}_n(u|F(x_0)) - \tilde{m}(u|F(x_0))| \\
&\quad \to \sup_{0\le u\le 1} |B_0(u|F(x_0))| \quad \text{in distribution.}
\end{aligned}$$

Here $\tilde{m}_n$ and $\tilde{m}$ are the processes pertaining to the "uniform" C_n and C. Observe, however, that for continuous $\tilde{m}(\cdot|F(x_0))$

$$\sup_{0\le u\le 1} |B_0(u|F(x_0))| = \sqrt{\int K^2(u)\,du}\ \sup_{0\le u\le 1} |B_0^*(u)|$$

in distribution, where B_0^* is a standard Brownian bridge on $[0,1]$. As a consequence, we see that the Kolmogorov–Smirnov test statistic leads to large-sample distribution-free tests and confidence bands for $m(\cdot|x_0)$.

2. A proper conditional empirical process. As mentioned earlier, m_n is not a proper distribution function. Alternatively, we might consider the function

$$m_n^*(\mathbf{y}|x_0) = \frac{\sum_{i=1}^{n} 1_{(-\infty,\mathbf{y}]}Y_i \cdot K\left(\frac{F_n(x_0)-F_n(X_i)}{a_n}\right)}{\sum_{i=1}^{n} K\left(\frac{F_n(x_0)-F_n(X_i)}{a_n}\right)}$$

a proper distribution function. Observe that

$$m_n^*(\mathbf{y}|x_0) = m_n(\mathbf{y}|x_0)/f_n(x_0),$$

where

$$f_n(x_0) = (na_n)^{-1}\sum_{i=1}^{n} K\left(\frac{F_n(x_0)-F_n(X_i)}{a_n}\right).$$

In other words, $f_n(x_0) = m_n(x_0)$ with $Y_i \equiv 1$. Since for such a Y one has $m(x_0) = 1$ and $\operatorname{var}(Y|X = x_0) = 0$ we obtain that

$$(na_n)^{1/2}[f_n(x_0) - 1] \to 0 \quad \text{in probability.}$$

It follows that under the smoothness assumptions of the theorem

$$(na_n)^{1/2}[m_n^*(\mathbf{y}|x_0) - \bar{m}_n(\mathbf{y}|x_0)] = (na_n)^{1/2}[m_n(\mathbf{y}|x_0) - \bar{m}_n(\mathbf{y}|x_0)] + o_{\mathbb{P}}(1) \quad \text{uniformly in } \mathbf{y}.$$

We thus see that m_n^* fulfills the same invariance principle as m_n.

3. Conditional quantiles. When $d = 1$, i.e., when Y is real-valued, the process m_n^* has an inverse or quantile function

$$m_n^{*-1}(u|x_0) = \inf\{y \in \mathbb{R} : m_n^*(y|x_0) \ge u\}, \qquad 0 < u < 1.$$

This is scheduled for estimating the u quantile of $m(\cdot|x_0)$. In this section we derive the limit distribution of

$$Q_n(u) \equiv (na_n)^{1/2}\left[m_n^{*-1}(u|x_0) - m^{-1}(u|x_0)\right], \qquad 0 < u < 1 \text{ fixed.}$$

For such an u, write $y_u = m^{-1}(u|x_0)$.

THEOREM 3. *Under the assumptions of the corollary, if* $m'(y_u|x_0) = (\partial/\partial y)m(y|x_0) > 0$ *at* $y = y_u$ *and G is continuous we have for almost all* x_0

$$Q_n(u) \to N(0, \sigma_u^2) \quad \textit{in distribution},$$

where

$$\sigma_u^2 = u(1-u)\int K^2(x)\,dx/[m'(y_u|x_0)]^2.$$

PROOF. The method is the same as for showing asymptotic normality of (unconditional) quantiles. [See, e.g., Wretman (1978).] Compared with Wretman's proof, we use C-tightness of m_n^* rather than Chebyshev's inequality (because of the heavy dependence of the summands) to show that when

$$W_n^* = (na_n)^{1/2}[m_n^*(y_u|x_0) - m(y_u|x_0)]$$

and

$$W_n = (na_n)^{1/2}\left[m_n^*\left(y_u + y/(na_n)^{1/2}|x_0\right) - m\left(y_u + y/(na_n)^{1/2}|x_0\right)\right],$$

then $W_n^* - W_n \to 0$ in probability. The theorem then immediately follows from asymptotic normality of W_n^* and continuity of the standard normal distribution function. See Wretman (1978) for details. □

4. Lemmas and proofs. Put

$$\beta_n(\mathbf{y}) \equiv \beta_n(\mathbf{y}|x_0) = (na_n)^{1/2}[m_n(\mathbf{y}|x_0) - \overline{m}_n(\mathbf{y}|x_0)].$$

We shall prove the theorem by showing that:

(i) the finite-dimensional distributions of β_n converge to those of B_0.

(ii) $\{\beta_n\colon n \ge 1\}$ is uniformly C-tight, i.e., for each $\varepsilon > 0$ and every $\rho > 0$ there exist some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$

$$\mathbb{P}\left(\sup_{\|\mathbf{y}_1 - \mathbf{y}_2\| \le \delta} |\beta_n(\mathbf{y}_1) - \beta_n(\mathbf{y}_2)| \ge \rho\right) \le \varepsilon.$$

Observe that $\beta_n(\mathbf{0}) = 0$ for each $n \in \mathbb{N}$. As to (i) we have the following

LEMMA 4. *Under the assumptions of the theorem, the finite-dimensional distributions of* β_n *converge to those of* B_0.

PROOF. Follows at once from the theorem in Stute (1984b) upon applying the Cramér–Wold device. □

To prove tightness we shall have to rest on some bounds for the oscillation modulus of multivariate empirical processes. For this, let $a(1), \dots, a(1+d)$ be some positive constants (which may depend on n) and put $\mathbf{a} = (a(1), \dots, a(1+d))$. For $\mathbf{x} = (x_1, \dots, x_{1+d}) \le \mathbf{y} = (y_1, \dots, y_{1+d})$ componentwise denote with $I_{\mathbf{x},\mathbf{y}} = \prod_{i=1}^{1+d}(x_i, y_i]$ the pertinent rectangle in R^{1+d}. In Stute (1984a) we derived finite sample upper bounds and almost sure limit results for maximal deviations of the empirical process

$$\alpha_n(x_1, \ldots, x_{1+d}) := n^{1/2}[H_n(x_1, \ldots, x_{1+d}) - H(x_1, \ldots, x_{1+d})]$$

over small rectangles. To be specific let

$$\omega_n(\mathbf{a}) := \sup\{|\alpha_n(I_{\mathbf{x},\mathbf{y}})|: |y_i - x_i| \le a(i) \text{ for } 1 \le i \le 1 + d\}$$

denote the oscillation modulus of α_n. Then it was shown in Theorem 1.7 of that paper that under certain growth assumptions on $a(1), \ldots, a(1 + d)$

$$(1) \qquad \mathbb{P}(\omega_n(\mathbf{a}) > s) \le C_1\Big[\min_{1 \le i \le 1+d} a(i)\Big]^{-(1+d)} \exp[-C_2 s^2/\min a(i)]$$

for some C_1, C_2 not depending on **a**, s, n, or H.

To study conditional empirical processes at a point $x_0 \in \mathbb{R}$, we shall have to restrict ourselves to rectangles $I_{\mathbf{x},\mathbf{y}}$ for which $x_1 \le x_0 \le y_1$. Write

$$\omega_n(\mathbf{a}; x_0) := \sup\{|\alpha_n(I_{\mathbf{x},\mathbf{y}})|: |y_i - x_i| \le a(i) \text{ for } 1 \le i \le 1 + d \text{ and } x_1 \le x_0 \le y_1\}$$

and put (with μ denoting the distribution pertaining to H)

$$\gamma(\mathbf{a}; \mathbf{x}_0) := \sup\{\mu(I_{\mathbf{x},\mathbf{y}})\}$$

with the supremum extended over the class of rectangles appearing in $\omega_n(\mathbf{a}; x_0)$. To motivate Lemma 5 below, we should like to mention that (1) had been derived by bounding $\omega_n(\mathbf{a})$ from above by the maximal deviation of α_n over a finite number of small rectangles $I_1, \ldots, I_m$ forming a partition of $[0,1]^{1+d}$, with the length of each side being of the order $\min_{1 \le i \le 1+d} a(i)$. Hence $m \sim [\min_{1 \le i \le 1+d} a(i)]^{-(1+d)}$. After that, an appropriate maximal inequality together with a standard Bernstein exponential bound applied to $\alpha_n|I_j$, $1 \le j \le m$, then yielded the desired bound (1). In the case of $\omega_n(\mathbf{a}; x_0)$, since $x_1 \le x_0 \le y_1$ for all rectangles in question, it suffices to partition the coordinate space $\prod_{i=2}^{1+d}[0,1]$ into small rectangles with each side having length of order $\min_{2 \le i \le 1+d} a(i)$. From this observation it is likely to obtain a bound for $\omega_n(\mathbf{a}; x_0)$ similar to (1), but with $[\min_{1 \le i \le 1+d} a(i)]^{-(1+d)}$ replaced by the smaller factor $[\min_{2 \le i \le 1+d} a(i)]^{-d}$. As remarked after Theorem 1.7 in Stute (1984a), the bound (1) may be improved if some further information on H is available, e.g., if H has a bounded Lebesgue density. In fact, the denominator $\min_{1 \le i \le 1+d} a(i)$ in the exponential factor occurs when applying the Bernstein bound by observing that for each $I_{\mathbf{x},\mathbf{y}}$ with $|y_i - x_i| \le a(i)$ we have $\mu(I_{\mathbf{x},\mathbf{y}}) \le \text{const}\, x\prod_{i=1}^{1+d} a(i)$. Noting that, by definition, $\gamma(\mathbf{a}; x_0)$ is a general upper bound for $\mu(I_{\mathbf{x},\mathbf{y}})$ we thus obtain:

LEMMA 5. *Suppose that H has uniform marginals. Then there exist constants $C_1, C_2 > 0$ (not depending on s, n, **a**, or H) such that*

$$(2) \qquad \mathbb{P}(\omega_n(\mathbf{a}; x_0) > s) \le C_1\Big[\min_{2 \le i \le 1+d} a(i)\Big]^{-d} \exp[-C_2 s^2/\gamma(\mathbf{a}; x_0)],$$

provided that $2 \le s\sqrt{n}$ and $C_3\gamma(\mathbf{a}; x_0) \ge s/\sqrt{n}$, C_3 finite.

We shall apply Lemma 5 to vectors $\mathbf{a} = (a(1), \dots, a(1 + d))$, where $a(1) = a_n \to 0$ at appropriate rates and $\min_{2 \le i \le 1+d} a(i) = \delta > 0$ is small but fixed.

LEMMA 6. $\{\beta_n\colon n \ge 1\}$ *is uniformly C-tight.*

PROOF. Write

$$\beta_n(\mathbf{y}|x_0) = \sqrt{na_n}\,[m_n(\mathbf{y}|x_0) - m_n^*(\mathbf{y}|x_0)] + \sqrt{na_n}\,[m_n^*(\mathbf{y}|x_0) - \overline{m}_n(\mathbf{y}|x_0)]$$
$$\equiv \beta_{n1}(\mathbf{y}|x_0) + \beta_{n2}(\mathbf{y}|x_0),$$

where

$$m_n^*(\mathbf{y}|x_0) = a_n^{-1}\int 1_{[0,\mathbf{y}]}(\mathbf{u})K\left(\frac{F_n(x_0) - F_n(x)}{a_n}\right)H(dx, d\mathbf{u}).$$

We show that both β_{n1} and β_{n2}, $n \ge 1$, are uniformly C-tight. As to β_{n1}, we have, upon integrating by parts,

$$m_n(\mathbf{y}|x_0) - m_n^*(\mathbf{y}|x_0) = a_n^{-1}[H_n(1,\mathbf{y}) - H(1,\mathbf{y})]\,K\left(\frac{F_n(x_0) - 1}{a_n}\right)$$
$$-a_n^{-1}\int [H_n(x,\mathbf{y}) - H(x,\mathbf{y})]\,K_n(dx)$$

with

$$K_n(x) = K\left(\frac{F_n(x_0) - F_n(x)}{a_n}\right).$$

Since $F_n(x_0) \to x_0$ $(0 < x_0 < 1)$ with probability one, $a_n \to 0$ and K has finite support, the first summand is zero with probability one for all $n \ge n_0(\omega)$, say, not depending on $\mathbf{y}$. Similarly, for $n \ge n_1(\omega)$

$$\int [H_n(x,\mathbf{y}) - H(x,\mathbf{y})]\,K_n(dx)$$
$$= \int [H_n(x,\mathbf{y}) - H(x,\mathbf{y}) - H_n(x_0,\mathbf{y}) + H(x_0,\mathbf{y})]\,K_n(dx).$$

Assume $K = 0$ outside $(-1, 1)$ w.l.o.g., i.e., the last integral remains unchanged when restricting the domain of integration to those x's for which $|F_n(x_0) - F_n(x)| \le a_n$. For given $\varepsilon > 0$ the Dvoretzky–Kiefer–Wolfowitz (1956) bound entails that for some finite (large) C_3 one has, up to an event of probability less than or equal to ε, that

$$|F(x_0) - F(x)| \le a_n + C_3 n^{-1/2} \le C_4 a_n$$

whenever $|F_n(x_0) - F_n(x)| \le a_n$. Denoting with $\|K\|$ the total variation of K we thus obtain for all large n, neglecting an event of probability $\le \varepsilon$, that

$$\sup_{\|\mathbf{y}_1 - \mathbf{y}_2\| \le \delta} |\beta_{n1}(\mathbf{y}_1|x_0) - \beta_{n1}(\mathbf{y}_2|x_0)| \le a_n^{-1/2}\|K\| \sum_{i=1}^{d} \omega_n(C_4 a_n, 1, \dots, \delta, 1, \dots, 1; x_0).$$

To bound the last sum, we may apply Lemma 5 with $s = \rho\sqrt{a_n}$ by observing that $\gamma(C_4 a_n, 1, \dots, \delta, \dots, 1; x_0) \geq C_4 a_n \delta$, so that the growth conditions are satisfied for at least all large n. For such an n

$$\mathbb{P}\Big(\omega_n(C_4 a_n, 1, \dots, \delta, 1, \dots, 1; x_0) > \rho\sqrt{a_n}\Big) \leq C_1 \delta^{-d} \exp\big[-C_2 \rho^2 a_n/\gamma\big].$$

By (A), $\gamma = o(a_n/\ln \delta^{-1})$ as $\delta \to 0$ and $n \to \infty$. Hence the last exponential bound can be made arbitrarily small for all $\delta \leq \delta_0$ and $n \geq n_0$, say. This proves tightness of β_{n1}.

As to β_{n2}, write

$$\begin{aligned} m_n^*(\mathbf{y}|x_0) &= a_n^{-1} \int 1_{[\mathbf{0},\mathbf{y}]}(\mathbf{u}) K\left(\frac{F(x_0) - F(x)}{a_n}\right) H(dx, d\mathbf{u}) \\ &\quad + a_n^{-2} \int 1_{[\mathbf{0},\mathbf{y}]}(\mathbf{u}) [F_n(x_0) - F_n(x) - F(x_0) + F(x_0)] \\ &\quad \times K'\left(\frac{F(x_0) - F(x)}{a_n}\right) H(dx, d\mathbf{u}) \\ &\quad + a_n^{-3} \int 1_{[\mathbf{0},\mathbf{y}]}(u) [F_n(x_0) - F_n(x) - F(x_0) + F(x)]^2 \\ &\quad \times K''(\Delta)/2 H(dx, d\mathbf{u}) \\ &= \overline{m}_n(\mathbf{y}|x_0) + I_2(\mathbf{y}, n) + I_3(\mathbf{y}, n) \end{aligned}$$

with Δ between $a_n^{-1}[F_n(x_0) - F_n(x)]$ and $a_n^{-1}[F(x_0) - F(x)]$. Similar to Lemma 1 of Stute (1984b) we get that $(na_n)^{1/2} I_3(y, n) \to 0$ in probability uniformly in $\mathbf{y}$. Thus to prove the lemma it remains to show that

$$\left\{\sqrt{na_n}\, I_2(\cdot, n) \colon n \geq 1\right\} \quad \text{is uniformly } C\text{-tight.}$$

With $\alpha_n(x) = n^{1/2}[F_n(x) - x]$, $0 \leq x \leq 1$, we have

$$\begin{aligned} \sqrt{na_n}\, I_2(\mathbf{y}, n) &= a_n^{-3/2} \int m(\mathbf{y}|x) [\alpha_n(x_0) - \alpha_n(x)] K'\left(\frac{x_0 - x}{a_n}\right) dx \\ &= a_n^{-3/2} \int [m(\mathbf{y}|x) - m(\mathbf{y}|x_0)] [\alpha_n(x_0) - \alpha_n(x)] K'\left(\frac{x_0 - x}{a_n}\right) dx \\ &\quad + a_n^{-3/2} m(\mathbf{y}|x_0) \int [\alpha_n(x_0) - \alpha_n(x)] K'\left(\frac{x_0 - x}{a_n}\right) dx. \end{aligned}$$

Use the same arguments as in the proof of Lemma 3 in Stute (1984b) to show that the first summand converges to zero in probability uniformly in $\mathbf{y}$ whenever $m(\cdot|x)$ is equicontinuous in a neighborhood of x_0. Finally, for large n,

$$\begin{aligned} & a_n^{-3/2} m(\mathbf{y}|x_0) \int [\alpha_n(x_0) - \alpha_n(x)] K'\left(\frac{x_0 - x}{a_n}\right) dx \\ &= -a_n^{-1/2} m(\mathbf{y}|x_0) \int K\left(\frac{x_0 - x}{a_n}\right) \alpha_n(dx). \end{aligned}$$

Since $a_n^{-1/2}\int K[(x_0 - x)/a_n]\alpha_n(dx)$ has a normal limit distribution and is hence stochastically bounded, and since $m(\cdot|x_0)$ is (uniformly) continuous, this proves tightness of β_{n2}. □

5. Concluding remark. It is possible to extend the results of this paper to multivariate X. We found it useful, however, to separate the univariate from the general case. In fact, regarding the distribution of X, our processes m_n (resp. m_n^*) turned out to be distribution-free. For multivariate X, the transformations involved lead to processes with underlying uniform marginals, but otherwise depending on the (joint) distribution of X.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

COLLOMB, G. (1981). Estimation non-paramétrique de la regression: Revue bibliographique. *Internat. Statist. Rev.* **49** 75–93.

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.

NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

STUTE, W. (1984a). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379.

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

WRETMAN, J. (1978). A simple derivation of the asymptotic distribution of a sample quantile. *Scand. J. Statist.* **5** 123–124.

MATHEMATICAL INSTITUTE
UNIVERSITY OF GIESSEN
ARNDTSTRAßE 2
D-6300 GIESSEN
WEST GERMANY