The Annals of Probability
1984, Vol. 12, No. 2, 361–379

THE OSCILLATION BEHAVIOR OF EMPIRICAL PROCESSES: THE MULTIVARIATE CASE

BY WINFRIED STUTE

University of Siegen[1]

We derive sharp finite sample estimates and exact almost sure limit results for local deviations of multivariate empirical processes. These are useful for obtaining, e.g., exact convergence rates of multivariate kernel density estimators. It is also indicated how local properties of multivariate empirical processes may be used to study various problems in nonparametric multivariate analysis.

0. Introduction. In this paper we derive sharp finite sample tail estimates and exact almost sure asymptotic results on local deviations of empirical processes in higher dimension. These correspond to those of Stute (1982a) for dimension one.

One major difficulty in dealing with multivariate empirical processes comes from the fact that the parameter set is no longer linearly ordered. It turns out that a detailed study of such processes must involve their exact path structure and distributional character. In Kiefer (1961) a certain conditioning technique using the Markovian structure of empirical distribution functions has been successfully applied to derive exponential bounds for the maximal (global) deviation between the empirical and the true distribution function. Here an appropriately modified method is used to derive bounds for the local deviation at a point and for the oscillation modulus (Section 1). The induction argument is based on a sharp exponential bound derived in Stute (1982a). Some asymptotic results are stated in Section 2.

As we have shown in our previous paper, an application of such local results easily yields tightness of the empirical process, both in the Skorokhod topology and under sup-norm metrics. We also obtained weak conditions for the convergence of the so-called quantile process, and gave a straightforward proof of Kiefer's uniform Bahadur representation of sample quantiles. Finally (see also Stute, 1982b) local properties turn out to be crucial for deriving exact rates of convergence for kernel density estimators, histograms and nearest neighbor estimators. In Section 3 of this paper, similar results are obtained for multivariate kernel density estimators. In Section 4, we discuss possible applications in various fields of nonparametric multivariate analysis. Some further applications are indicated in Section 5. Proofs are given in Section 6.

In the following let $\xi, \xi_1, \xi_2, \cdots$ denote an independent sample in $\mathbb{R}^d$, $d \geq 1$,

Received May 1982; revised October 1983.

[1] Now at the University of Giessen.

AMS 1980 *subject classifications*. Primary 60F15, 60G17; secondary 62G05.

Key words and phrases. Multivariate empirical process, oscillation modulus, copula function, kernel density estimators, copula process, conditional distribution function, regression function.

with common distribution function (d.f.)

$$H(\mathbf{t}) = \mathbb{P}(\xi \le \mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^d,$$

defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let H_n be the empirical d.f. of the sample $\xi_1, \cdots, \xi_n$, and let

$$\alpha_n(\mathbf{t}) \equiv n^{1/2}(H_n(\mathbf{t}) - H(\mathbf{t})), \quad \mathbf{t} \in \mathbb{R}^d,$$

denote the corresponding empirical process. Similarly, for a rectangle I, put

$$\alpha_n(I) \equiv n^{1/2}(\mu_n(I) - \mu(I)),$$

where μ_n and μ are the probability measures pertaining to H_n and H, respectively.

Now, it will be convenient to deal with a specific representation of α_n. For this, write $H(\mathbf{t}) \equiv H(t_1, \cdots, t_d) = C(F_1(t_1), \cdots, F_d(t_d))$, where $F_1, \cdots, F_d$ are the marginals of H and C is the pertaining copula (or dependence) function (see, e.g., Schweizer and Wolff, 1981). In particular C is a d.f. on the unit cube in $\mathbb{R}^d$ with uniform marginals. Let $\eta_i = (\eta_i^1, \cdots, \eta_i^d)$, $i = 1, 2, \cdots$ be an independent sample with d.f. C. Put $\xi_i = (F_1^{-1}(\eta_i^1), \cdots, F_d^{-1}(\eta_i^d))$, $i = 1, 2, \cdots$. Then $\xi_1, \xi_2, \cdots$ are independent with d.f. H. Denote with C_n the empirical d.f. of $\eta_1, \cdots, \eta_n$. It is easy to see that $H_n(t_1, \cdots, t_d) = C_n(F_1(t_1), \cdots, F_d(t_d))$, so that

$$\alpha_n(\mathbf{t}) = n^{1/2}[C_n(F_1(t_1), \cdots, F_d(t_d)) - C(F_1(t_1), \cdots, F_d(t_d))]. \tag{0.1}$$

Since C has uniform marginals it will therefore be sufficient to consider the case when $F_1 = \cdots = F_d = U$, the uniform distribution on the unit interval. We shall mainly be concerned with the process $\{\alpha_n(I): I \in \mathscr{I}\}$, where $\mathscr{I}$ is some class of (small) rectangles in the unit cube, possibly depending on n. For notational convenience, write $I_{\mathbf{x},\mathbf{y}} = \prod_{1\le i\le d}(x_i, y_i]$ whenever $\mathbf{x} = (x_1, \cdots, x_d) \le \mathbf{y} = (y_1, \cdots, y_d)$ componentwise. Furthermore, put $\| I_{\mathbf{x},\mathbf{y}} \| = \max_{1\le i\le d}(y_i - x_i)$. For $0 \le a \le b < \infty$, $\mathscr{I}(a, b)$ is the class of rectangles $I = I_{\mathbf{x},\mathbf{y}}$ with $\max_{1\le i\le d}(y_i - x_i) \le b$ and $\min_{1\le i\le d}(y_i - x_i) \ge$ a. Thus, $I \in \mathscr{I}(0, b)$ if and only if $\| I \| \le b$. In our local investigations the vector $\mathbf{a} = (a_1, \cdots, a_d) \in [0, 1]^d$ always serves as the right upper corner of a (small) rectangle with left lower corner $\mathbf{0} = (0, \cdots, 0) \in [0, 1]^d$.

1. Local deviations. In this section we shall state some finite sample bounds for local increments of α_n. The basic inequality (1.1) below is obtained by induction on d. The case $d = 1$ has been treated in Stute (1982a). See also Shorack and Wellner (1982).

LEMMA 1.1. *Let G be a continuous* d.f. *on the real line with inverse $G^{-1}(p) = \inf\{t \in \mathbb{R}: G(t) \ge p\}$, $0 \le p \le 1$. Then, for each $0 < \delta < 1$ and $0 < a < \delta/4$, if $8 \le (s\delta)^2$,*

$$\mathbb{P}(\sup_t \alpha_n(t) > s\sqrt{a}) \le 2\, \mathbb{P}(\bar{\alpha}_n(a) > s(1 - \delta)\sqrt{a}),$$

with the supremum extended over all t between $G^{-1}(0)$ and $G^{-1}(a)$.

Here α_n and $\bar{\alpha}_n$ denote the empirical processes pertaining to samples of size n with d.f.'s G and U, respectively.

PROOF. It follows from Lemma 2.3 in Stute (1982a) upon using the representation $\alpha_n(t) = \bar{\alpha}_n(G(t))$. □

A similar bound will now be stated for a general d. As for Kiefer's (1961) tail estimates for $\| H_n - H \| = \sup_{\mathbf{t}} | H_n(\mathbf{t}) - H(\mathbf{t}) |$, the method of proof exploits the Markovian structure of H_n (Lemma 6.1). Though, the resulting estimates are completely different in our case, thus reflecting the local character of the setup. In the following, H will be assumed to be continuous with uniform marginals. In particular, the support of H is contained in the unit cube. Fix some $0 < a_1, \cdots, a_d < 1/2$ such that $H(\mathbf{a}) \equiv H(a_1, \cdots, a_d) < \delta/4$, where $\delta < 1/2$.

LEMMA 1.2. *Under the above assumptions there exists some constant* $C = C(\delta) < \infty$ *such that*

$$(1.1) \qquad \mathbb{P}(\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}}\alpha_n(\mathbf{t}) > s\sqrt{H(\mathbf{a})}) \le C\,\mathbb{P}(\alpha_n(\mathbf{a}) \ge s(1-2\delta)^d\sqrt{H(\mathbf{a})})$$

for all $s > 0$ *with* $2 \le s\sqrt{nH(\mathbf{a})}$ *and* $32 \le (s\delta(1-2\delta))^2$.

Since in applications δ will be a small positive number and $C = C(\delta)$ is finite, (1.1) tells us that essentially nothing is lost when replacing the supremum by $\alpha_n(\mathbf{a})$. The proof of Lemma 1.2 will be given in the last section of this paper. An explicit bound for the right hand side of (1.1) is obtained from a standard Bernstein-inequality for the Binomial random variable $nH_n(\mathbf{a}) = \eta$. Though there are somewhat sharper bounds available, the present one is appropriate for most purposes.

LEMMA 1.3. *Given* $0 < \delta < 1$, *there exists some (small)* $x_\delta > 0$ *such that for a* $\mathrm{Bin}(n, p)$*-random variable* η

$$\mathbb{P}(|\eta - np| \ge z) \le 2\exp[-(1-\delta)z^2/(2np)],$$

provided that $0 \le z \le npx_\delta$.

Lemma 1.3 provides a type of exponential bound which is usually needed for proving LIL results. It is also valid with $2np(1-p)$ rather than $2np$. Since we shall only deal with small p, the factor $(1-p)$ may be thought of as being absorbed by $(1-\delta)$.

A bound similar to (1.1) may also be obtained for the probability that $\inf_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}}\alpha_n(\mathbf{t}) < -s\sqrt{H(\mathbf{a})}$. Together with Lemma 1.3 we therefore get the following estimate.

THEOREM 1.4. *For each* $0 < \delta < 1/2$ *there exists some finite* $C = C(\delta)$ *such that* *for all* $0 < a_1, \cdots, a_d < \frac{1}{2}$ *with* $H(a_1, \cdots, a_d) < \delta/4$ *and every* $0 < s \le x_\delta \sqrt{nH(\mathbf{a})}$, *one has*

(1.2) $\mathbb{P}(\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}} |\alpha_n(\mathbf{t})| > s\sqrt{H(\mathbf{a})}) \le C \exp[-(1-\delta)(1-2\delta)^{2d}s^2/2]$.

PROOF. Of course, it remains to show (1.2) for all $s \ge c = c(\delta) > 0$, say, since for $s \le c$, (1.2) is trivially true for C large enough. Now, since $s \le x_\delta \sqrt{nH(\mathbf{a})}$, we may find some $c = c(\delta) > 0$ such that the growth conditions in Lemma 1.2 are satisfied for $s \ge c$. With $z = s(1-2\delta)^d \sqrt{nH(\mathbf{a})}$ the result therefore follows from (1.1) and Lemma 1.3. □

In most applications both $\mathbf{a} = \mathbf{a}(n)$ and $s = s_n$ will depend on n in such a way that $H(\mathbf{a}(n)) \to 0$ and $nH(\mathbf{a}(n)) \to \infty$, with $s_n = o(\sqrt{nH(\mathbf{a}(n))})$. Given $\delta > 0$, let $\varepsilon < \frac{1}{2}$ be such that $1-\delta \le (1-\varepsilon)(1-2\varepsilon)^{2d}$. Then (1.2) is applicable for ε, at least for all large n. Hence

(1.3) $\mathbb{P}(\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}(n)} |\alpha_n(\mathbf{t})| > s_n\sqrt{H(\mathbf{a}(n))}) \le C(\delta)\exp[-(1-\delta)s_n^2/2]$.

As in Stute (1982a), the estimates (1.2) and (1.3) can be used to study the oscillation modulus of α_n defined by

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

For this one has to observe that the bound (1.2) also holds if instead of $\alpha_n(\mathbf{t})$, $\mathbf{0} \le \mathbf{t} \le \mathbf{a}$, one considers $\alpha_n(I_{\mathbf{x},\mathbf{t}})$, $\mathbf{x} \le \mathbf{t} \le \mathbf{x} + \mathbf{a}$, with $H(\mathbf{a})$ replaced by $\mu(I_{\mathbf{x},\mathbf{x}+\mathbf{a}})$. Application to every $\mathbf{x}$ of an appropriate finite grid of the unit cube then yields an upper bound for the tails of $\omega_n(\mathbf{a})$. We shall first state an estimate for the important case when H has a smooth density f w.r.t. Lebesgue measure on the unit cube which is bounded below away from zero.

THEOREM 1.5. *Suppose that H has a continuous density f on $[0, 1]^d$, with $f \ge m > 0$, and uniform marginals. Then for given $0 < \delta < \frac{1}{2}$, there exist some (large) finite $C = C(\delta)$ and some (small) $y_\delta > 0$ such that for all small enough $0 < a_1, \cdots, a_d$ and every $0 < s \le y_\delta\sqrt{n}\,\min_{1\le i\le d} a_i^d$*

(1.4) $\mathbb{P}(\omega_n(\mathbf{a}) > s\sqrt{a_1 \cdots a_d \sup_x f(x)}) \le C\,[\min_{1\le i\le d} a_i]^{-d}\exp[-(1-\delta)s^2/2]$.

The boundedness condition $f \ge m > 0$ is important if one is interested in bounds for ω_n which are associated with the tails of a normal distribution (cf. Lemma 1.3). If one admits more general (but less tractable) bounds for binomial tails (see Bennett, 1962), the condition $f \ge m > 0$ may be omitted.

It may happen that one is interested in fluctuations of α_n when restricted to a fixed subrectangle I_0 of the unit cube, namely

$$\omega_n(\mathbf{a}, I_0) \equiv \sup\{|\alpha_n(I_{\mathbf{x},\mathbf{y}})|: y_i - x_i \le a_i \text{ for } 1 \le i \le d,\ I_{\mathbf{x},\mathbf{y}} \subset I_0\}.$$

A slight modification of the proof of Theorem 1.5 then shows that (1.4) is still true with the supremum extended over all $\mathbf{x} \in I_0$. In this case only $f \mid I_0 \ge m > 0$ is needed.

As in Stute (1982a) and (1982b), local properties of empirical processes suggest themselves for an empirical study of the unknown density f of H. For this it is necessary to study local deviations of α_n on small but not too small rectangles. Recall that, for $0 \le a \le b < \infty$, $\mathscr{J}(a, b)$ denotes the class of rectangles

$$I_{\mathbf{x},\mathbf{y}} \text{ with } \max_{1\le i\le d}(y_i - x_i) \le b \text{ and } \min_{1\le 1\le d}(y_i - x_i) \ge a.$$

THEOREM 1.6. *Under the conditions of Theorem* 1.5, *then, for given* $0 < \delta < 1$ *and* $0 < c_1 \le c_2 < \infty$, *there exists some* $y(\delta, c_1, c_2) > 0$ *such that for all small* $a > 0$ *and* $0 < s \le y\sqrt{na^d}$

$$\mathbb{P}\left(\sup_{I\in\mathscr{J}(c_1a,c_2a)} \frac{|\alpha_n(I)|}{\sqrt{\mu(I)}} > s\right) \le C\, a^{-d}\exp[-(1-\delta)s^2/2], \tag{1.5}$$

where $C < \infty$ *depends only on* c_1, c_2, δ *and* d.

So far we have stated our main results under the additional assumption that H has uniform marginals. For the general (smooth) case the corresponding estimates are easily obtained from (0.1). For example, (1.2) yields the same bound for $\mathbb{P}(\sup_{\mathbf{t}} |\alpha_n(t)| > s\sqrt{H(\mathbf{a})})$, provided that $F_i(a_i) \le \frac{1}{2}$, $1 \le i \le d$, and where the supremum extends over all $\mathbf{t}$ with $t_i \le a_i$, $1 \le i \le d$.

Also, (1.5) is valid in more general situations. For example, if F_i, $1 \le i \le d$, have derivatives F_i' such that (on their support) $0 < M_1 \le F_i' \le M_2 < \infty$, then $c_1a \le y_i - x_i \le c_2a$, $1 \le i \le d$, implies $c_1M_1a \le F_i(y_i) - F_i(x_i) \le c_2M_2a$, $1 \le i \le d$. Hence (0.1) and (1.5) applied to c_1M_1 and c_2M_2 rather than c_1 and c_2 yield the same bound for an arbitrary (smooth) $H = C(F_1, \cdots, F_d)$ (possibly with a somewhat larger constant). Note also that if H has a (smooth) positive density f and positive marginal densities F_i', $1 \le i \le d$, the copula function C has the density

$$k(z_1, \cdots, z_d) = f(F_1^{-1}(z_1), \cdots, F_d^{-1}(z_d))/[\textstyle\prod_{1\le i\le d} F_i'(F_i^{-1}(z_i))],$$

$0 < z_i < 1$, $1 \le i \le d$. Clearly, the boundedness condition $k \ge m > 0$ is too restrictive for most applications, if one admits uniformly continuous H with unbounded support. Observe, however, that as remarked before, for the transformed empirical process a version of (1.4) and hence of (1.5) is also valid when restricted to a fixed subrectangle of the unit cube. For the non-transformed empirical process, this entails that given a rectangle $I_0 \subset \mathbb{R}^d$, a version of (1.5) is valid (with the supremum restricted to all $I \subset I_0$), if $k \mid (F_1, \cdots, F_d)(I_0) \ge m > 0$. Under some mild smoothness assumptions on f, this is easily verified for each bounded I_0. For example, consider the case when, on $I_0 = \prod_{1\le i\le d}(x_i, y_i)$, $f \mid I_0$ and $F_i' \mid (x_i, y_i)$ are bounded away from zero and infinity.

As to (1.4) one has to be more careful. Take $d = 2$ for simplicity and write $F \equiv F_1$, $G \equiv F_2$, $a \equiv a_1$ and $b \equiv a_2$. Assume F', $G' \le M_2$. For $I = (x_1, y_1] \times (x_2, y_2]$ we have $F(y_1) - F(x_1) \le M_2a$ and $G(y_2) - G(x_2) \le M_2b$ whenever $y_1 - x_1 \le a$, $y_2 - x_2 \le b$. For $k \in \mathbb{N}$ put $J_i = [i/k, (i+2)/k]$, $i = 0, 1, \cdots, k-2$. If M_2a, $M_2b \le 1/k$, then $F(x_1)$, $F(y_1) \in J_j$ and $G(x_2)$, $G(y_2) \in J_l$ for some j and l. Let $q < 1$

be close to one and assume that

$$(*)\qquad \begin{aligned} &\inf_{z\in J_j} F'(F^{-1}(z)) \ge q \sup_{z\in J_j} F'(F^{-1}(z)),\\ &\inf_{z\in J_l} G'(G^{-1}(z)) \ge q \sup_{z\in J_l} G'(G^{-1}(z)). \end{aligned}$$

Condition $(*)$ cannot be satisfied for j, $l = 0$ or $k - 2$, if the density quantile functions $F'(F^{-1}(z))$ and $G'(G^{-1}(z))$ tend to zero as $z \to 0$ or 1. It will, however, be satisfied under suitable regularity assumptions on F and G, for all J_j and J_l contained in some $[\varepsilon, 1 - \varepsilon]$, say, when k is large. Now, using (0.1), let c_n be defined by the equation $\alpha_n(t_1, t_2) = c_n(F(t_1), G(t_2))$. Obviously the event $\{|\alpha_n(I)| > s\sqrt{ab}\cdot\sup_{\mathbf{x}\in I_0} f(\mathbf{x})\}$ then implies the event

$$\{\bar\omega_n(\bar a, \bar b, J_j \times J_l) > s\sqrt{\bar a\bar b}\, \sup_{\mathbf{x}\in I_0} f(\mathbf{x})/[\sup_{z\in J_j} F'(F^{-1}(z)) \cdot \sup_{z\in J_l} G'(G^{-1}(z))]\}$$

$$\subset \{\bar\omega_n(\bar a, \bar b, J_j \times J_l) > sq\sqrt{\bar a\bar b}\, \sup_{(z_1,z_2)\in J_j\times J_l} k(z_1, z_2)\},$$

where $\bar\omega_n$ is the oscillation modulus of c_n, and $\bar a = a \sup_{z\in J_j} F'(F^{-1}(z))$, $\bar b = b \sup_{z\in J_l} G'(G^{-1}(z))$. Now apply (1.4) to bound the probability of the last event. For bounded I_0's summation over appropriate j's and l's then gives the bound (1.4) for $\mathbb{P}(\omega_n(a, b, I_0) > s\sqrt{ab}\, \sup_{\mathbf{x}\in I_0} f(\mathbf{x}))$.

So far the results of this section have been stated for a "standardized" α_n. For example, (1.1) yields an upper bound for the tails of $\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}} \alpha_n(\mathbf{t})/\sqrt{H(\mathbf{a})}$. This is necessary if one is interested in exact rates of convergence for $\omega_n(\mathbf{a}(n)) \to 0$ as $n \to \infty$, and similarly for the results of Section 3. If one is merely interested in suitable upper bounds it suffices to bound $\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}} \alpha_n(\mathbf{t})$ and $\omega_n(\mathbf{a})$. For such an estimate the conditions may be substantially weakened. To be precise, replace s in Lemma 1.2 by $s/\sqrt{H(\mathbf{a})}$ (assume $H(\mathbf{a}) > 0$ since otherwise the resulting estimate becomes trivial). The condition $2 \le s\sqrt{nH(\mathbf{a})}$ then reduces to $2 \le s\sqrt{n}$. Since $H(\mathbf{a}) \le \min_{1\le i\le d} a_i$ we therefore get

$$(1.1)^* \qquad \mathbb{P}(\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{a}} \alpha_n(\mathbf{t}) > s) \le C\, \mathbb{P}(\alpha_n(\mathbf{a}) > s(1 - 2\delta)^d),$$

provided that $2 \le s\sqrt{n}$, $32H(\mathbf{a}) \le (s\delta(1 - 2\delta))^2$ and $\min_{1\le i\le d} a_i < \delta/4$. Instead of Lemma 1.3 the following version of a Bernstein-inequality is more appropriate to bound the right-hand side of (1.1)*:

$$(1.6) \qquad \mathbb{P}(|\eta - np| \ge \sqrt{n}s) \le 2\exp[-s^2/(2p + (2s/3\sqrt{n}))], \quad s > 0.$$

Using (1.1)* and (1.6) rather than (1.1) and Lemma 1.3, we then obtain the following result.

THEOREM 1.7. *Suppose that H has uniform marginals. Then there exist constants $C_1, C_2 > 0$ (not depending on s, n, $a_1, \cdots, a_d$ or H) such that*

$$(1.7) \qquad \mathbb{P}(\omega_n(\mathbf{a}) > s) \le C_1[\min_{1\le i\le d} a_i]^{-d} \exp[-C_2 s^2/\min_{1\le i\le d} a_i],$$

provided that $2 \le s\sqrt{n}$ and $C_3 \min a_i \ge s/\sqrt{n}$, C_3 finite.

When H has a bounded Lebesgue-density, the exponential term may be improved to $\exp[-C_2 s^2/\prod a_i]$, provided that $C_3 \prod a_i \ge s/\sqrt{n}$.

From Theorem 1.7 it is now trivial, via Borel-Cantelli, to obtain almost sure upper bounds for $\omega_n(\mathbf{a}(n))$ under various growth conditions on $\mathbf{a}(n) = (a_{1n}, \cdots, a_{dn})$.

2. Asymptotic results. In Stute (1982a) the one-dimensional analogue of (1.4) has been used to study the oscillation modulus of α_n as $a = a_n \downarrow 0$, $n \to \infty$. For an almost sure limit result, the following mild growth conditions for a_n were needed:

$$\text{(i) } na_n \uparrow \infty \quad \text{(ii) } \ln a_n^{-1} = o(na_n) \quad \text{(iii) } \ln a_n^{-1}/\ln\ln n \to \infty.$$

Such sequences were called bandsequences. In dimension d (i)–(iii) have to be assumed for a_n^d rather than a_n.

THEOREM 2.1. *Suppose that H has a continuous density f on the unit cube with $f \geq m > 0$, and uniform marginals. Then, if $(a_n^d)_n$ is a bandsequence, we have with probability one*

$$\lim_{n\to\infty} \omega_n(a_n, \cdots, a_n)/\sqrt{2a_n^d \ln a_n^{-d}} = \sqrt{\sup_{\mathbf{x}} f(\mathbf{x})}. \tag{2.1}$$

PROOF. This follows from Theorem 1.5 in much the same way as Theorems 2.14 and 2.15 in Stute (1982a) followed from Lemma 2.4 there. In fact, that $d = 1$ was not very essential there. In particular, the poissonization argument is also valid in higher dimension. That a_n has to be replaced by a_n^d comes from (1.4) and the fact that for general d the growth of $\omega_n(a_n, \cdots, a_n)$ is related to the maximal deviation of α_n on small pairwise disjoint rectangles $I_1, \cdots, I_r$ where $\| I_j \| = O(a_n)$, $j = 1, \cdots, r$ and r is of order a_n^{-d}. □

Similarly, from (1.5) we get the following result.

THEOREM 2.2. *Under the assumptions of Theorem* 2.1 *we have for all* $0 < c_1 \leq c_2 < \infty$ *with probability one*

$$\lim_{n\to\infty} \sup_{I \in \mathscr{I}(c_1 a_n, c_2 a_n)} \frac{| \alpha_n(I) |}{\sqrt{2\mu(I) \ln a_n^{-d}}} = 1. \tag{2.2}$$

This is the d-dimensional analogue of Theorem 2.12 in Stute (1982a).

If $F_1, \cdots, F_d$ are not necessarily uniform on [0, 1], Theorems 2.1 and 2.2 take on the following form.

THEOREM 2.3. *Suppose that on $I_0 = \prod_{1\leq i\leq d}(x_i, y_i)$, H admits a uniformly continuous density f such that $0 < m \leq f \mid I_0 \leq M < \infty$ and $0 < M_1 \leq F_i' \leq M_2 < \infty$, $1 \leq i \leq d$. Then for all $0 < c_1 \leq c_2 < \infty$, we have with probability one*

$$\lim_{n\to\infty} \omega_n(a_n, \cdots, a_n, I_0)/\sqrt{2a_n^d \ln a_n^{-d}} = \sqrt{\sup_{\mathbf{x}\in I_0} f(\mathbf{x})} \tag{2.1$'$}$$

$$\lim_{n\to\infty} \sup_{I \in \mathscr{I}(c_1 a_n, c_2 a_n), I \subset I_0} \frac{| \alpha_n(I) |}{\sqrt{2\mu(I) \ln a_n^{-d}}} = 1, \tag{2.2$'$}$$

provided that $(a_n^d)_n$ is a bandsequence.

3. Kernel density estimators. Suppose that H has a nonspecified (Lebesgue -) density f. There is an extensive literature on nonparametric methods for estimating f. A survey of available methods is contained in Wertz (1978). Perhaps the best studied estimator is the so-called kernel density estimator

$$f_n(\mathbf{t}) = a_n^{-d} \int K\left(\frac{\mathbf{t} - \mathbf{x}}{a_n}\right) H_n(d\mathbf{x}), \quad \mathbf{t} \in \mathbb{R}^d,$$

introduced for $d = 1$ by Rosenblatt (1956) and Parzen (1962). Here K is a probability kernel on $\mathbb{R}^d$ and $(a_n)_n$ is a bandwidth-sequence tending to zero as $n \to \infty$. Nadaraya (1965) was the first to establish rates for the almost sure uniform convergence of f_n to f. He always considers kernel functions of bounded variation, thus making integration by parts possible. After that an application of the Dvoretzky-Kiefer-Wolfowitz (1956) exponential bound for the (global) deviation between the empirical and the hypothetical d.f. then yields the desired rate. A more detailed investigation of the large sample behavior of kernel type estimates may be found in Révész (1978), who uses strong approximation results for empirical processes. See also Révész (1982) and Silverman (1978). In the multivariate case ($d > 1$) kernel density estimators have been first investigated by Cacoullos (1964). Rüschendorf (1977) extended Nadaraya's (1965) work, using Kiefer's (1961) estimate instead of the D-K-W bound. It should be obvious, however, that such global estimates are not appropriate to get exact rates for local type estimates. As to Révész's (1978) technique, the existing results on strong approximation of multivariate empirical processes (cf. Csörgő and Révész, 1975, Philipp and Pinzur, 1980) do not yield the same bounds as obtained by the direct local approach described below. To this end take $c_1 = 1 = c_2$ in (2.2)$'$. Then each I has the form $I = I_{\mathbf{t}-\mathbf{a}_n/2,\mathbf{t}+\mathbf{a}_n/2}$ for some $\mathbf{t}$. Put

$$f_n(\mathbf{t}) = a_n^{-d}\mu_n(I_{\mathbf{t}-\mathbf{a}_n/2,\mathbf{t}+\mathbf{a}_n/2}) \text{ and } \bar{f}_n(\mathbf{t}) = a_n^{-d}\mu(I_{\mathbf{t}-\mathbf{a}_n/2,\mathbf{t}+\mathbf{a}_n/2}),$$

which is the expectation of $f_n(\mathbf{t})$. Note that f_n is the d-dimensional analogue of the naive kernel density estimator, since

$$f_n(\mathbf{t}) = a_n^{-d} \int K\left(\frac{\mathbf{t} - \mathbf{x}}{a_n}\right) H_n(d\mathbf{x}), \tag{3.1}$$

where (in obvious notation) $K = 1_{[-1/2,1/2)}$. Under the smoothness assumptions of Theorem 2.3 a version of (2.2)$'$ is still true with $\mu(I)$ replaced by $\lambda(I)f(\mathbf{t})$, where λ is Lebesgue measure on $\mathbb{R}^d$. Thus (2.2)$'$ implies

$$\lim_{n\to\infty}\sup_{\mathbf{t}\in I_0} \sqrt{na_n^d}\, \frac{|f_n(\mathbf{t}) - \bar{f}_n(\mathbf{t})|}{\sqrt{2f(\mathbf{t}) \ln a_n^{-d}}} = 1 \tag{3.2}$$

with probability one.

As in Stute (1982b) the relation 3.2 may be extended to an arbitrary simple kernel function. The general form of the right-hand side of (3.2) then becomes $(\int K^2(\mathbf{x})\, d\mathbf{x})^{1/2}$. In a second step the same may be obtained for an arbitrary smooth kernel function of bounded support, which is of finite variation. For this,

K has to be approximated by a suitable simple kernel function K^0 to which the general form of (3.2) applies. Furthermore, it must be shown that the normalized error between the two resulting estimators f_n and f_n^0 can be made arbitrarily small if K^0 gets close to K. In this part of the proof an integration by parts argument and the fact that K has bounded support will be needed to bound the deviation between f_n and f_n^0 by the oscillation modulus of α_n. To make integration by parts possible, K has to be of finite variation in the sense of Hardy and Krause (see, e.g., Hlawka, 1961, and Hobson, 1927). In particular, this is fulfilled if K has finite support and bounded partial derivatives of order two. In summary, we get the following result.

THEOREM 3.1. *Let K be a continuous kernel function with bounded support and finite variation. Suppose that on $I_0 = \prod (x_i, y_i)$, $H = C(F_1, \cdots, F_d)$ has a uniformly continuous density f such that $f \mid I_0$ and $F_i' \mid (x_i, y_i)$, $1 \le i \le d$, are bounded away from zero and infinity. Then with probability one*

$$\text{(3.3)} \qquad \lim_{n\to\infty} \sup_{\mathbf{t}\in I_0} \sqrt{\frac{na_n^d}{2 \ln a_n^{-d}}} \frac{|f_n(\mathbf{t}) - \bar{f}_n(\mathbf{t})|}{\sqrt{f(\mathbf{t})}} = \left(\int K^2(\mathbf{x})\, d\mathbf{x}\right)^{1/2},$$

provided that $(a_n^d)_n$ is a bandsequence.

In density estimation theory, the sequence $(a_n)_n$ is usually called the sequence of window-widths.

As in Stute (1982b), certain extensions of (3.3) are possible, such as for kernel functions with unbounded support, but with certain growth properties at infinity.

From (3.3) it is now possible to derive exact error terms for the deviation between f_n and f. To get explicit bounds for the bias $\bar{f}_n - f$, some smoothness assumptions will be needed for f. In principle, to require a certain type of smoothness is a matter of taste, according to what class of K's is to be considered and what rate of convergence one has in mind. As a rule, bounding the bias is a trivial analytical task. Simple Taylor expansion will usually do the job.

In this paper we shall only be concerned with nonnegative kernels integrating to one (probability kernel). Then the following smoothness assumption on f is appropriate:

$$\text{(3.4)} \qquad f(\mathbf{t} + \mathbf{h}) = f(\mathbf{t}) + \{f'(\mathbf{t})\}^T\mathbf{h} + \tfrac{1}{2}\, \mathbf{h}^T\{f''(\mathbf{t})\}\mathbf{h} + o(\mathbf{h}^T\mathbf{h})$$

uniformly in a $\mathbf{t}$-neighbourhood of I_0, as $\mathbf{h} \to 0$. Here

$$f'(\mathbf{t}) = \left(\frac{\partial f}{\partial x_i}, 1 \le i \le d\right)_{\mathbf{x}=\mathbf{t}}, \quad f''(\mathbf{t}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{\mathbf{x}=\mathbf{t}},$$

and T denotes transposition. Under the structural assumptions of Theorem 3.1 we get

$$a_n^{-2}[\bar{f}_n(\mathbf{t}) - f(\mathbf{t})] = \frac{1}{2} \int K(\mathbf{y})\mathbf{y}^T\{f''(\mathbf{t})\}\mathbf{y}\, d\mathbf{y} + o(1)$$

uniformly in $\mathbf{t} \in I_0$, provided that K is symmetric. Hence, from (3.3),

$$\lim_{n\to\infty}\sup_{\mathbf{t}\in I_0} \sqrt{\frac{na_n^d}{2 \ln a_n^{-d}}} \frac{|f_n(\mathbf{t}) - f(\mathbf{t})|}{\sqrt{f(\mathbf{t})}} = \left(\int K^2(\mathbf{x})\, d\mathbf{x}\right)^{1/2},$$

whenever $a_n^2 = o((\ln a_n^{-d}/na_n^d)^{1/2})$, i.e. when $na_n^{4+d}/\ln a_n^{-d} \to 0$. On the other hand, the stochastic component $f_n - \bar{f}_n$ is negligible, if $na_n^{4+d}/\ln a_n^{-d} \to \infty$. The optimal rate of convergence is obtained if a_n^2 and $(\ln a_n^{-d}/na_n^d)^{1/2}$ are of the same order. This is achieved if we set $a_n = \text{const} \times (\ln n/n)^{1/(4+d)}$, where the constant depends on f (and f'') and is hence unknown (see Stute, 1982b, for $d = 1$). It should be worthwhile making some comments on the growth conditions (ii) and (iii) (with a_n replaced by a_n^d). Now, (3.3) trivially implies

$$\sup_{\mathbf{t}\in I_0} |f_n(\mathbf{t}) - \bar{f}_n(\mathbf{t})| = O(((\ln a_n^{-d})/na_n^d)^{1/2})$$

with probability one, and, as we have seen, similarly for the deviation between f_n and f. Of course, to obtain consistency of f_n, the last term should be $o(1)$. Hence (ii) is not only a sufficient but also a desirable growth condition on a_n (respectively a_n^d). Condition (iii) prevents a_n^d from being too large. Roughly, it has to be slightly smaller than $1/\ln n$. Such a condition is necessary to obtain equality in (2.1), (2.2) and (3.3), rather than an upper bound. See Stute (1982a) for details.

4. Applications in multivariate nonparametric analysis. In this section we shall indicate how Theorem 1.7 may be applied to various problems in nonparametric multivariate analysis. To this end, recall $H = C(F_1, \cdots, F_d)$, with C denoting the copula function of H. For $F_1, \cdots, F_d$ continuous, we have

(4.1) $$C(u_1, \cdots, u_d) = H(F_1^{-1}(u_1), \cdots, F_d^{-1}(u_d)), \quad 0 \le u_i \le 1 \quad \text{for} \quad i = 1, \cdots, d.$$

It is easily seen that the (empirical) copula function C_n pertaining to H_n is also given, in analogy to (4.1), by

$$C_n(u_1, \cdots, u_d) = H_n(F_{1n}^{-1}(u_1), \cdots, F_{dn}^{-1}(u_d)),$$

with F_{in} denoting the ith marginal d.f. of H_n. Putting $(u_1, \cdots, u_d) = (i_1/n, \cdots, i_d/n)$, $1 \le i_1, \cdots, i_d \le n$, we see that C_n is a certain multiparameter rank type process (see Rüschendorf, 1976). It is further known (cf, e.g., Schweizer and Wolff, 1981), that

$$H(t_1, \cdots, t_d) = F_1(t_1) \cdots F_d(t_d) \quad \text{for all real} \quad t_1, \cdots, t_d$$

if and only if

$$C(u_1, \cdots, u_d) = u_1 \cdots u_d \quad \text{for all} \quad 0 \le u_i \le 1.$$

It is therefore reasonable to base tests on multivariate independence on the process

$$C_n(u_1, \cdots, u_d) - u_1 \cdots u_d, \quad 0 \le u_i \le 1, \quad 1 \le i \le d.$$

The collection of statistics, which can be defined through C_n, includes Kolmogorov-Smirnov and Cramér-von Mises type statistics as well as Spearman and Kendall type rank correlation coefficients. For nondegenerate limit results one has to study the so-called (standardized) copula process

$$c_n(u_1, \cdots, u_d) = n^{1/2}[C_n(u_1, \cdots, u_d) - C(u_1, \cdots, u_d)], \quad 0 \le u_i \le 1.$$

The main difficulty in dealing with C_n comes from the fact that C_n is equal to H_n evaluated at the random points $(F_{1n}^{-1}(u_1), \cdots, F_{dn}^{-1}(u_d))$. It turns out, however, that the distribution of C_n does not depend on $F_1, \cdots, F_d$, so that again we may assume w.l.o.g. $F_i = U$, the uniform distribution on the unit interval. Observe that in this case

$$\| F_{in}^{-1} - F_i^{-1} \| = \sup_u | F_{in}^{-1}(u) - F_i^{-1}(u) | = \sup_t | F_{in}(t) - F_i(t) |,$$

$1 \le i \le d$. Hence, by the Chung-Smirnov LIL for empirical d.f.'s, we obtain $\| F_{in}^{-1} - F_i^{-1} \| \to 0$ as $O((\ln\ln n/n)^{1/2})$, with probability one. We may therefore apply Theorem 1.7 and the Borel-Cantelli lemma (with appropriate $a = a_n$ and $s = s_n$) to get bounds for the convergence

$$\sup_{u_1, \cdots, u_d} | \alpha_n(F_{1n}^{-1}(u_1), \cdots, F_{dn}^{-1}(u_d)) - \alpha_n(u_1, \cdots, u_d) | \to 0,$$

from which it is possible to eliminate the random effect in the argument of H_n. In particular, it may be shown that under appropriate smoothness assumptions on C, c_n has the representation

$$\begin{aligned} c_n(u_1, \cdots, u_d) &= \bar{\alpha}_n(u_1, \cdots, u_d) - \sum_{i=1}^d \frac{\partial}{\partial u_i} C(u_1, \cdots, u_d)\bar{\alpha}_n(1, \cdots, u_i, \cdots, 1) \\ &\quad + R_n(u_1, \cdots, u_d) \end{aligned}$$

where $\bar{\alpha}_n$ is the empirical process pertaining to $C(=H$ if $F_1 = \cdots = F_d = U)$ and the error term satisfies

$$\sup | R_n(u_1, \cdots, u_d) | = O(n^{-1/4}(\ln n)^{1/2}(\ln\ln n)^{1/4}) \tag{4.2}$$

as $n \to \infty$, with probability one. Clearly, such a representation easily yields Strassen-Finkelstein and Donsker type invariance principles for c_n (see Gaenssler and Stute, 1979).

The copula function also enters, via its partial derivatives, into the representation of the conditional d.f. of ξ^2 given ξ^1, where $\xi = (\xi^1, \xi^2)$ and ξ^i are $\mathbb{R}^{d_i}$-valued random vectors $(i = 1, 2)$ with $d_1 + d_2 = d$:

$$\mathbb{P}(\xi^2 \le (t_{d_1+1}, \cdots, t_d) \mid \xi^1 = (t_1, \cdots, t_{d_1})) = C_0(F_1(t_1), \cdots, F_d(t_d)),$$

where

$$\begin{aligned} C_0(u_1, \cdots, u_d) &= \int_0^{u_{d_1+1}} \cdots \int_0^{u_d} k(u_1, \cdots, u_{d_1}, z_1, \cdots, z_{d_2})\, dz_1 \cdots dz_{d_2} \\ &= \frac{\partial^{d_1}}{\partial u_1 \cdots \partial u_{d_1}} C(u_1, \cdots, u_d). \end{aligned}$$

In his interesting discussion of Stone's (1977) paper on nonparametric regression, Professor Parzen proposed estimation of the conditional d.f. by the function (in our terms)

$$(t_1, \cdots, t_d) \to \tilde{C}_n(F_{1n}(t_1), \cdots, F_{dn}(t_d)).$$

He, and also Professor Stone in his reply, admit that "many theorems remain to be proved" and "much work seems required" to justify "the efficiency of this procedure". It will be indicated below that as for (4.2), local results for α_n will at least be helpful to give more insight into the structure of $\tilde{C}_n(F_{1n}, \cdots, F_{dn})$. For example, consider the naive kernel estimate of C_0. In this case $\tilde{C}_n(F_{1n}(t_1), \cdots, F_{dn}(t_d))$ is equal to the (standardized) number of sample points $\xi_j = (\xi_j^1, \xi_j^2)$, $1 \le j \le n$, for which $\xi_j^2 \le (t_{d_1+1}, \cdots, t_d)$, subject to the condition (where a_n denotes a preassigned bandwidth)

$$\left(F_{1n}^{-1}\left(F_{1n}(t_1) - \frac{a_n}{2}\right), \cdots, F_{d_1n}^{-1}\left(F_{d_1n}(t_{d_1}) - \frac{a_n}{2}\right)\right)$$

$$< \xi_j^1 \le \left(F_{1n}^{-1}\left(F_{1n}(t_1) + \frac{a_n}{2}\right), \cdots, F_{d_1n}^{-1}\left(F_{d_1n}(t_{d_1}) + \frac{a_n}{2}\right)\right),$$

or equivalently (with ξ_{ij} denoting the ith component of ξ_j)

$$F_{in}(t_i) - \frac{a_n}{2} < F_{in}(\xi_{ij}) \le F_{in}(t_i) + \frac{a_n}{2} \quad \text{for all} \quad 1 \le i \le d_1. \tag{4.3}$$

The last condition expresses the fact that subject to $\xi_j^2 \le (t_{d_1+1}, \cdots, t_d)$ one has to count those ξ_j^1's which are close to $(t_1, \cdots, t_{d_1})$ in the sense of (4.3). This notion of closeness, which uses an a_n-neighborhood principle, is defined through the sample and does not use the particular (Euclidean) metric on the sample space.

Our bounds on local deviations of multivariate empirical (and copula) processes turn out to be the key tool in order to get a representation of the process $\tilde{C}_n(F_{1n}, \cdots, F_{dn})$. From this it will be possible to get exact rates of convergence and weak limit results for the process

$$(t_1, \cdots, t_d) \to [\tilde{C}_n(F_{1n}(t_1), \cdots, F_{dn}(t_d)) - C_0(F_1(t_1), \cdots, F_d(t_d))],$$

and to determine the optimal rate at which the "bandwidths" a_n should converge to zero. Since these problems are more statistical in nature, a detailed study of these processes will be contained in a separate paper.

Moreover, once knowing the structure of the process $\tilde{C}_n(F_{1n}, \cdots, F_{dn})$ it will also be possible to get Bahadur-Kiefer type representations of the pertaining quantiles (when $d_2 = 1$). More generally, one may study the conditional M-estimate θ_n (of location), being the root of the equation

$$\int \psi(t - \theta)\tilde{C}_n(F_{1n}(t_1), \cdots, F_{d-1,n}(t_{d-1}), F_{dn}(dt)) = 0,$$

where ψ is a given score-function. In particular, putting $\psi(z) = z$, we get

$$\theta_n = \int t\, \tilde{C}_n(F_{1n}(t_1), \cdots, F_{d-1,n}(t_{d-1}), F_{dn}(dt))$$

as an estimate of the regression function $\mathbb{E}(\xi^2 \mid \xi^1 = (t_1, \cdots, t_{d-1}))$.

5. Further applications. As in our previous paper Stute (1982a), a straightforward application of Theorem 1.7 yields tightness and hence, in view of the convergence of the finite dimensional distributions, the invariance principle for empirical processes in higher dimensions. Furthermore, local results are useful to prove the convergence of α_n and c_n in sup-norm metrics. In particular, the latter is important for proving, e.g., weak limit results for rank correlation statistics $\int\int J \circ C_n du_1 \cdots du_d$ with unbounded score functions J. As one further application we mention that local type results for empirical processes are useful for handling various problems in nonparametric sequential analysis. For example, they play an important role for obtaining efficient fixed-width confidence intervals for the unknown density. See Stute (1983) for the univariate case. Finally, local results are useful for improving the existing strong approximation results for multivariate empirical processes (cf. Csörgő and Révész, 1975, Philipp and Pinzur, 1980, and Csörgő, 1979).

6. Proofs. We shall first consider the case $d = 2$. Lemma 1.2 will then be an easy consequence of Lemmas 6.1 and 6.2 below. Write $a = a_1$ and $b = a_2$. Now, take some $0 < q < 1$ to be specified later on, and let the grid $\{t_j\}_{j=0,\cdots,j_0}$ in $[0, a]$ be defined by $t_j = a(1 - q^j)$, $j = 0, 1, \cdots, j_0$, and $t_{j_0+1} = a$, where $j_0 \in \mathbb{N}$. Suppose that $\mathbf{x}' = (x_1', x_2') \in [0, a) \times [0, b)$, and let j be such that $t_j \le x_1' < t_{j+1}$. In the conditioning argument to follow we shall be concerned with independent Bernoulli random variables each having probability of success $p = [H(t_{j+1}, x_2') - H(x_1', x_2')]/(1 - H(x_1', b))$. Observe that since $p \le (t_{j+1} - t_j)/(1 - t_{j+1})$, we have

$$p \le \frac{aq^j - aq^{j+1}}{1 - a + aq^{j+1}} \le \frac{1-q}{q} \quad \text{for} \quad j < j_0$$

and

$$p \le \frac{aq^{j_0}}{1-a} \le \frac{4q^{j_0}}{3} \quad \text{for} \quad j = j_0.$$

Hence choosing $q < 1$ so large that $(1 - q)/q \le \delta$ and then making j_0 so large that $4q^{j_0}/3 \le \delta$, we obtain a grid $\{t_j\}$ such that $p \le \delta$ for each $\mathbf{x}' \in [0, a) \times [0, b)$. Furthermore, j_0 depends only on δ but not on a, b, n, s or H.

That p can be made arbitrarily small by an appropriate partitioning of the interval $[0, a]$ makes things easier compared with global estimates of Kiefer (1961), where his Lemma 5-m and Lemma 6-m had to be included in order to handle the case when x_1' is close to one.

Now, similar to Kiefer (1961), define, for $j = 1, \cdots, j_0 + 1$

$$V_j = \{(x_1, x_2): 0 \le x_2 \le b,\ t_{j-1} \le x_1 \le t_j\}$$

$$W_j = \{(x_1, x_2): 0 \le x_2 \le b,\ x_1 = t_j\},$$

and let, with $r = s\sqrt{H(a, b)}$,

$$B_j(s) = \{\sup_{\mathbf{x}\in V_j} \alpha_n(\mathbf{x}) \ge r\}$$

$$C_j(s) = \{\sup_{\mathbf{x}\in W_j} \alpha_n(\mathbf{x}) \ge r\}.$$

Put $C_0(s) = \varnothing$.

LEMMA 6.1. *Suppose that* $2 \le r\sqrt{n}$ *and* $32 \le (\delta s)^2$. *Then*

$$\mathbb{P}(C_j(s(1-2\delta)) \mid \bar{C}_{j-1}(s), B_j(s)) \ge 1 - (\delta s/4)^{-2} \ge 1/2, \quad j = 1, \cdots, j_0 + 1.$$

PROOF. We have to show the first inequality, the second being a trivial consequence of our second growth assumption. Now, if $B_j(s)$ occurs, there exists a smallest x_1' in $t_{j-1} \le x_1 \le t_j$ for which $\alpha_n(x_1', x_2) \ge r$, for some x_2, $0 \le x_2 \le b$. Let x_2' be the smallest of such x_2, thus defining uniquely a random vector $X' = (X_1', X_2')$. We shall show that for each possible value $\mathbf{x}^0 = (x_1^0, x_2^0)$ of X'

$$\mathbb{P}(\alpha_n(t_j, x_2^0) \ge r(1-2\delta) \mid \bar{C}_{j-1}(s), B_j(s), X' = \mathbf{x}^0) \ge 1 - (\delta s/4)^{-2},$$

proving the Lemma. Since $\bar{C}_{j-1}(s)$ occurs, $x_1^0 > t_{j-1}$. By continuity of the marginals,

$$nH_n(x_1^0, b) \le nH(x_1^0, b) + rn^{1/2} + 1$$

with probability one. Let N denote the number of sample points not contained in $[0, x_1^0] \times [0, b]$. Then

$$N \ge n(1 - H(x_1^0, b)) - rn^{1/2} - 1 \equiv M.$$

If $M \le 0$, we have $rn^{-1/2} \ge 1 - H(x_1^0, b) - n^{-1} \ge 1 - t_j - n^{-1}$ and thus

$$\begin{aligned} H_n(t_j, x_1^0) - H(t_j, x_2^0) &\ge H_n(x_1^0, x_2^0) - H(x_1^0, x_2^0) + H(x_1^0, x_2^0) - H(t_j, x_2^0) \\ &\ge rn^{-1/2} - [H(t_j, b) - H(t_{j-1}, b)] \\ &\ge rn^{-1/2} - \frac{H(t_j, b) - H(t_{j-1}, b)}{1 - t_j}(n^{-1} + rn^{-1/2}) \\ &\ge rn^{-1/2}(1-\delta) - \delta n^{-1} \ge rn^{-1/2}(1-2\delta), \end{aligned}$$

by choice of q and j_0, and the fact that $1 \le rn^{1/2}$. If $M > 0$ and hence $N > 0$, let Q denote the event

$$Q = \{\alpha_n(t_j, x_2^0) - \alpha_n(x_1^0, x_2^0) \ge -2\delta r\}.$$

For the proof of the lemma it suffices to show that for each $v \ge M$

$$\mathbb{P}(Q \mid \bar{C}_{j-1}(s), B_j(s), X' = \mathbf{x}^0, N = v) \ge 1 - (\delta s/4)^{-2}. \tag{6.1}$$

Since Q occurs if and only if

$$\frac{n[H_n(t_j, x_2^0) - H_n(x_1^0, x_2^0)] - vp}{\sqrt{vp(1-p)}} \geq \frac{-2n^{1/2}\delta r + n[H(t_j, x_2^0) - H(x_1^0, x_2^0)] - vp}{\sqrt{vp(1-p)}} = y$$

and since the conditioning event in (6.1) depends only on $\{H_n(x_1, x_2): x_1 \leq x_1^0, x_2 \leq b\}$, the conditional probability of Q is equal to the unconditional probability that a standardized Binomial random variable exceeds y. The relation $M \leq v \leq n$ implies

$$y \leq \frac{n^{1/2}r(-2\delta + p) + p}{\sqrt{vp(1-p)}} \leq \frac{-n^{1/2}r\delta/2}{\sqrt{vp(1-p)}} \leq \frac{-r\delta/2}{\sqrt{p}} \leq -\delta s/4.$$

Relation (6.1) now easily follows from Chebyshev's inequality. □

In the following Lemma we shall derive an upper exponential bound for $\mathbb{P}(C_j(s))$.

LEMMA 6.2. *Suppose that $H(a, b) < \delta/4$ and $8 \leq (s\delta(1 - 2\delta))^2$ are satisfied. Then*

$$\mathbb{P}(C_j(s)) \leq 2\mathbb{P}(\alpha_n(a, b) \geq s(1-\delta)\sqrt{H(a, b)}).$$

Similarly, for $\mathbb{P}(C_j(s(1-2\delta)))$, i.e. for s replaced by $s(1-2\delta)$.

PROOF. Consider α_n restricted to $W_j \cup \{(x_1, x_2): x_1 \geq t_j, x_2 = b\} \equiv T$. Since the parameter set T is homeomorphic to a closed subinterval of the real line such that the corresponding rectangles $[0, \mathbf{x}]$, $\mathbf{x} \in T$, are linearly ordered, we may apply Lemma 1.1 with a replaced by $H(a, b)$ to get the desired result. □

Using the inequality

$$\mathbb{P}(B_j(s)) \leq \mathbb{P}(C_{j-1}(s)) + \frac{\mathbb{P}(C_j(s(1-2\delta)))}{\mathbb{P}(C_j(s(1-2\delta)) \mid \bar{C}_{j-1}(s), B_j(s))}, \quad j \geq 1,$$

we obtain from Lemma 6.1 and Lemma 6.2

$$\begin{aligned} \mathbb{P}(B_j(s)) &\leq \mathbb{P}(C_{j-1}(s)) + 2\,\mathbb{P}(C_j(s(1-2\delta))) \\ &\leq 2\,\mathbb{P}(\alpha_n(a, b) \geq s(1-\delta)\sqrt{H(a, b)}) \\ &\quad + 4\,\mathbb{P}(\alpha_n(a, b) \geq s(1-\delta)(1-2\delta)\sqrt{H(a, b)}). \end{aligned} \tag{6.2}$$

PROOF OF LEMMA 1.2. Since

$$\{\sup\nolimits_{0 \leq t \leq (a,b)} \alpha_n(t) > s\sqrt{H(a, b)}\} \subset \cup_{j=1}^{j_0+1} B_j(s),$$

the assertion follows from (6.2) and the fact that j_0 only depends on δ.

For a general $d \geq 2$ we have to set

$$V_j = \{(x_1, \cdots, x_d): 0 \leq x_i \leq a_i \text{ for } 2 \leq i \leq d,\ t_{j-1} \leq x_1 \leq t_j\}$$

$$W_j = \{(x_1, \cdots, x_d): 0 \leq x_i \leq a_i \text{ for } 2 \leq i \leq d,\ x_1 = t_j\}.$$

On $B_j(s)$, the random vector $X' = (X_1', \cdots, X_d')$ is defined in close analogy to the case $d = 2$, thus denoting, in a sense, the "smallest" point $\mathbf{x} = (x_1, \cdots, x_d)$ in the unit cube for which $\alpha_n(\mathbf{x}) \geq r$. A version of Lemma 6.1 also holds, with the same method of proof, for arbitrary d, while a corresponding Lemma 6.2 follows by induction on d. In summary, this proves Lemma 1.2 for a general d. □

Theorems 1.5—1.7 will be rather obvious from our fundamental inequality (1.1). For notational convenience we shall only treat the case $d = 2$. Write $a = a_1$ and $b = a_2$.

PROOF OF THEOREM 1.5. Let R be the smallest positive integer satisfying $\delta^2 \min(a, b)/144 > 1/R$. Write $I_{ij}(\mathbf{t}) = I_{(i/R, j/R),(i/R, j/R)+\mathbf{t}}$ for $\mathbf{t} \in [0, 1]^2$. To bound $\omega_n(a, b)$, observe that each $I = I_{\mathbf{x},\mathbf{y}}$, $y_1 - x_1 \leq a$, $y_2 - x_2 \leq b$, may be represented as $I = I_{ij}(\mathbf{t}) + I_2 + I_3 + I_4$ for some $\mathbf{t} \leq (a, b)$ and $0 \leq i, j \leq R - 1$, where I_2 and I_3 (possibly empty) are adjacent to $I_{ij}(\mathbf{t})$, the horizontal (vertical) side of I_2 (I_3) being less than or equal to $1/R$, the other being less than or equal to $b(a)$. The rectangle I_4 (if nonempty) has one point in common with the boundary of $I_{ij}(\mathbf{t})$, with both sides being less than or equal to $1/R$. Simple inspection now shows that

$$\begin{aligned}
\omega_n(a, b) \leq\ & \max_{0 \leq i,j \leq R-1} \sup_{\mathbf{0} \leq \mathbf{t} \leq (a,b)} |\alpha_n(I_{ij}(\mathbf{t}))| \\
&+ 2 \max_{0 \leq i,j \leq R-1} \sup_{\mathbf{0} \leq \tau \leq (1/R, b)} |\alpha_n(I_{ij}(\tau))| \\
&+ 2 \max_{0 \leq i,j \leq R-1} \sup_{\mathbf{0} \leq \tau \leq (a, 1/R)} |\alpha_n(I_{ij}(\tau))| \\
&+ 4 \max_{0 \leq i,j \leq R-1} \sup_{\mathbf{0} \leq \tau \leq 1/R} |\alpha_n(I_{ij}(\tau))|.
\end{aligned} \tag{6.3}$$

For given $0 < \delta < \frac{1}{2}$ the left-hand side of (1.4) is therefore less than or equal to

$$\begin{aligned}
&\textstyle\sum_{i,j=0}^{R-1} \mathbb{P}(\sup_{\mathbf{0} \leq \mathbf{t} \leq (a,b)} |\alpha_n(I_{ij}(\mathbf{t}))| > s\sqrt{\mu(I_{ij}(a, b))}/(1 + \delta)) \\
&\quad + \textstyle\sum_{i,j=0}^{R-1} \mathbb{P}(\sup_{\mathbf{0} \leq \tau \leq (1/R, b)} |\alpha_n(I_{ij}(\tau))| > s\delta\sqrt{ab \sup f}/6(1 + \delta)) \\
&\quad + \textstyle\sum_{i,j=0}^{R-1} \mathbb{P}(\sup_{\mathbf{0} \leq \tau \leq (a, 1/R)} |\alpha_n(I_{ij}(\tau))| > s\delta\sqrt{ab \sup f}/6(1 + \delta)) \\
&\quad + \textstyle\sum_{i,j=0}^{R-1} \mathbb{P}(\sup_{\mathbf{0} \leq \tau \leq \mathbf{1/R}} |\alpha_n(I_{ij}(\tau))| > s\delta\sqrt{ab \sup f}/12(1 + \delta)).
\end{aligned} \tag{6.4}$$

By choice of R,

$$\delta\sqrt{ab \sup f}/6 \geq \delta\sqrt{ab \sup f}/12 \geq \sqrt{\mu(I_{ij}(x, y))}$$

where $(x, y) = (a, 1/R)$, $(1/R, b)$ or $(1/R, 1/R)$. Let $y_\delta = \delta^2 x_\delta \sqrt{m}/288 > 0$. If $s \leq y_\delta \sqrt{n} \min(a^2, b^2)$, then for each (i, j)

$$\begin{aligned}
s &\leq \delta^2 x_\delta \sqrt{nm} \min(a^2, b^2)/288 \leq x_\delta \sqrt{nm}/2(R - 1) \leq x_\delta \sqrt{nm}/R \\
&\leq x_\delta \sqrt{n\mu(I_{ij}(x, y))} \leq x_\delta \sqrt{n\mu(I_{ij}(a, b))}.
\end{aligned}$$

We may therefore apply (1.2) to bound each of the above probabilities. Hence, for some finite C,

$$\mathbb{P}(\omega_n(a, b) > s\sqrt{ab}\sup f) \leq CR^2 \exp[-(1-\delta)(1-2\delta)^4 s^2/2(1+\delta)^2].$$

For small enough a and b, we may do the same reasoning with some $\delta' > 0$ such that the last exponent is less than $-(1-\delta)s^2/2$, where $\delta > 0$ is given. This proves the theorem. □

PROOF OF THEOREM 1.6. Fix some $0 < q = q(\delta) < 1$ to be specified later on, and let $m_0 \in \mathbb{N} \cup \{0\}$ be defined by the relation $q^{m_0+1}c_2 \leq c_1 \leq q^{m_0}c_2$. For $0 < \alpha \leq \beta$, $0 < \gamma \leq \rho$ and $I = (x_1, y_1] \times (x_2, y_2]$, write $I \in \mathscr{J}^0(\alpha, \beta; \gamma, \rho)$ iff $\alpha \leq y_1 - x_1 \leq \beta$ and $\gamma \leq y_2 - x_2 \leq \rho$. Since

$$\mathscr{J}(c_1 a, c_2 a) \subset \cup_{l,k=0}^{m_0} \mathscr{J}^0(q^{l+1}c_2 a, q^l c_2 a; q^{k+1}c_2 a, q^k c_2 a)$$

and m_0 depends only on δ, c_1 and c_2, it suffices to bound $\sup_I |\alpha_n(I)|/\sqrt{\mu(I)}$, when I ranges over one of the above $\mathscr{J}^0$'s. Now, by assumption, f is uniformly continuous with $0 < m \leq f \leq M < \infty$. From this it is easy to see that $\inf_{I_1} f/\sup_{I_2} f \geq q$ for each I_1 and every I_2, for which $\|I_1\|$, $\|I_2\|$ and the Hausdorff distance between I_1 and I_2 are sufficiently small. Now, fix l and k and let I be a member of the pertaining $\mathscr{J}^0$. Introduce R as in the proof of Theorem 1.5, but for $(q^l c_2 a, q^k c_2 a)$ rather than (a, b). Consider the partition $I = I_{ij}(\mathbf{t}) + I_2 + I_3 + I_4$, and take $K_{ij} = \{(z_1, z_2): i/R \leq z_1 \leq i/R + q^l c_2 a, j/R \leq z_2 \leq j/R + q^k c_2 a\}$, so that $I_{ij}(\mathbf{t}) \subset K_{ij}$. We have (with λ for Lebesgue measure)

$$\mu(I) \geq \inf_I f \cdot \lambda(I) \geq q^3 \sup_{K_{ij}} f \cdot (c_2 a)^2 q^{l+k} \geq q^3 \mu(K_{ij}).$$

This gives

$$\frac{|\alpha_n(I_{ij}(\mathbf{t}))|}{\sqrt{\mu(I)}} \leq q^{-3/2} \sup_{0 \leq \tau; (q^l c_2 a, q^k c_2 a)} |\alpha_n(\mathrm{I}_{ij}(\tau))|/\sqrt{\mu(K_{ij})}.$$

The rectangles I_2, I_3 and I_4 are treated similarly. Hence, as in the proof of Theorem 1.5, we get in summary

$$\mathbb{P}\left(\sup_{I \in \mathscr{J}(c_1 a, c_2 a)} \frac{|\alpha_n(I)|}{\sqrt{\mu(I)}} > s\right) \leq Ca^{-2} \exp[-(1-\delta)(1-2\delta)^4 q^3 s^2/2(1+\delta)^2].$$

Finally, for a given $0 < \delta < 1$, choose some small $\delta' > 0$ and some large $q < 1$ such that the last exponent is less than $-(1-\delta)s^2/2$. The same reasoning for δ' then yields the assertion of the theorem. □

PROOF OF THEOREM 1.7. Choose some $0 < \delta < 1/2$ and let R be as in the proof of Theorem 1.5. Since C_1 and C_2 are unspecified, we may assume w.l.o.g. that the second and third growth condition for (1.1)* are satisfied. Recall (6.3) and use (6.4), in a non-standardized form. Since $a, b > 1/R$ the bound (1.1)* may also be applied to the summands in line 2-4 of (6.4). The assertion now follows from (1.6) and the fact that the probability of success $p = \mu(I(x, y))$ is less than or equal to $\min(a, b)$. □

When H has a bounded Lebesgue density an improved exponential factor may be obtained by observing that the probability p is now of order $o(\prod a_i)$.

REFERENCES

[1] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[2] CACOULLOS, T. (1964). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 179–189.

[3] CSÖRGŐ, M. (1979). Strong approximation of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process. *J. Mult. Anal.* **9** 84–100.

[4] CSÖRGŐ, M. and RÉVÉSZ, P. (1975). A strong approximation for the multivariate empirical process. *Studia Sci. Math. Hung.* **10** 427–434.

[5] 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.

[6] GAENSSLER, P. and STUTE, W. (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab.* **7** 193–243.

[7] HLAWKA, E. (1961). Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali Mat. Pura Appl* **54** 325–333.

[8] HOBSON, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series (3nd ed.), Vol 1.* Cambridge University Press, Cambridge, Mass.

[9] KIEFER, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

[10] NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probab. Appl.* **10** 186–190.

[11] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

[12] PHILIPP, W. and PINZUR, L. (1980). Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. verw. Gebiete* **54** 1–13.

[13] RÉVÉSZ, P. (1978). A strong law of the empirical density function. *Period. Math. Hung.* **9** 317–324.

[14] RÉVÉSZ, P. (1982). On the increments of Wiener and related processes. *Ann. Probab.* **10** 613–622.

[15] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[16] RÜSCHENDORF, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4** 912–923.

[17] RÜSCHENDORF, L. (1977). Consistency of estimators for multivariate density functions and for the mode. *Sankhyā Ser. A* **39** 243–250.

[18] SCHWEIZER, B. and WOLFF, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.* **9** 879–885.

[19] SHORACK, G. and WELLNER, J. (1982). Limit theorems and inequalities for the uniform empirical process indexed by intervals. *Ann. Probab.* **10** 639–652.

[20] SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184.

[21] STONE, C. (and discussants) (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

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

[23] STUTE, W. (1982b). A law of the logarithm for kernel density estimators. *Ann. Probab.* **10** 414–422.

[24] STUTE, W. (1983). Sequential fixed-width confidence intervals for a nonparametric density function. *Z. Wahrsch. verw. Gebiete* **62** 113–123.
[25] WERTZ, W. (1978). *Statistical Density Estimation. A Survey.* Vandenhoeck and Ruprecht, Göttingen.

MATHEMATISCHES INSTITUT
UNIVERSITÄT GIESSEN
ARNDTSTRASSE 2
D-6300 GIESSEN
WEST GERMANY