The Annals of Probability
1982, Vol. 10, No. 1, 86–107

THE OSCILLATION BEHAVIOR OF EMPIRICAL PROCESSES

BY WINFRIED STUTE[1]

*University of Munich**

In this paper we study the local behavior of empirical processes for independent identically distributed random variables on the real line. The results are applied to get best rates of convergence for various types of density estimators as well as error estimates for the Bahadur representation of the quantile process obtained by Kiefer.

0. Introduction and Main Results. Let $\xi_1, \xi_2, \cdots$ be a sequence of independent, identically distributed (i.i.d.) random variables on the real line, defined on some probability space $(\Omega, \mathscr{A}, \mathbb{P})$. Denote with

$$F(t) := \mathbb{P}(\{\omega \in \Omega : \xi_1(\omega) \le t\}), \qquad t \in \mathbb{R},$$

the joint (continuous) underlying distribution function (d.f.). Suppose that F is partially or completely unknown. One problem which arises in this context is that of constructing, for each $n \in \mathbb{N}$, an empirical estimate $F_n = F_n(\xi_1, \cdots, \xi_n)$ of F. The empirical d.f. which has been vastly investigated in the literature is obtained by assigning equal mass $1/n$ to each of the observations, i.e.

$$F_n(t) := n^{-1} \sum_{i=1}^n 1_{(-\infty,t]} \circ \xi_i, \qquad t \in \mathbb{R},$$

where 1_B denotes the indicator function of $B \subset \mathbb{R}$. The article of Gaenssler and Stute (1979) reviews some of the most important properties of F_n, both in the finite and infinite sample case.

Suppose that $\bar{\xi}_1, \bar{\xi}_2, \cdots$ is an i.i.d. sample with uniform distribution on the unit interval. Let

$$F^{-1}(p) := \inf\{t \in \mathbb{R} : F(t) \ge p\}, \qquad 0 < p < 1,$$

denote the inverse function of F. It is then straightforward to see that $\xi_i := F^{-1}(\bar{\xi}_i)$, $i = 1, 2, \cdots$ is an i.i.d. sample with distribution function F. Hence in what follows we may and do assume that ξ_i has the above representation. Write $\bar{F}_n$ for the empirical d.f. of $\bar{\xi}_1, \cdots, \bar{\xi}_n$. Since $F^{-1}(p) \le t$ if and only if $p \le F(t)$ we obtain

$$F_n(t) = \bar{F}_n(F(t)), \qquad t \in \mathbb{R}. \tag{0.1}$$

The corresponding empirical process is defined by

$$\alpha_n(t) = n^{1/2}[F_n(F^{-1}(t)) - t], \qquad 0 < t < 1.$$

Note that $F(F^{-1}(t)) = t$ for all t if F is continuous. Thus, by (0.1),

$$\alpha_n(t) = n^{1/2}(\bar{F}_n(t) - t), \qquad 0 \le t \le 1,$$

where $\alpha_n(0) = 0 = \alpha_n(1)$. Hence under a continuity assumption on F, α_n is the empirical process pertaining to a uniform sample. In particular, it is distribution-free.

Received June 9, 1980; revised January, 1981.

[1] This research is part of the author's "Habilitationsschrift" written at the Mathematical Department of the University of Munich.

Now at the University of Siegen.

* *AMS* 1970 *subject classifications*: Primary 60F15, 60G17; secondary 62G05

Key words and phrases: Empirical process, oscillation modulus, quantile process, Bahadur representation, naive density estimator, histograms.

The now classical invariance principle of Donsker states that

(0.2) $$\alpha_n \to_{\mathscr{L}} B^\circ \quad \text{as} \quad n \to \infty,$$

where "$\mathscr{L}$" denotes convergence in distribution in the Skorokhod-space $D[0, 1]$ (cf. Billingsley (1968)). Here $B^\circ(t) = B(t) - tB(1)$, $0 \le t \le 1$, is a tied-down Brownian Motion (Brownian Bridge), a centered Gaussian process with continuous sample paths, $B^\circ(0) = 0 = B^\circ(1)$, and covariance

$$\operatorname{cov}(B^\circ(s), B^\circ(t)) = s(1-t) \quad \text{for} \quad 0 \le s \le t \le 1.$$

For the proof of (0.2) one has to show that

(*a*) α_n has asymptotically the same finite-dimensional distributions as B°,

(*b*) if $\mathscr{L}\{\alpha_n\}$ denotes the distribution of α_n, then $\{\mathscr{L}\{\alpha_n\} : n \in \mathbb{N}\}$ is relatively compact in the space of all probability measures on $D[0, 1]$.

While (*a*) is an immediate consequence of the multidimensional CLT, the proof of (*b*) may be finished in two steps. First one shows, using an Arzelà-Ascoli type argument for the space $D[0, 1]$, that (*b*) is implied by the following condition:

(0.3) For each $\varepsilon > 0$ and every $\eta > 0$ there exists some $a > 0$ such that for all large enough $n \in \mathbb{N}$

$$\mathbb{P}(\{\omega_n(a) \ge \varepsilon\}) \le \eta,$$

where

$$\omega_n(a) = \sup_{|t-s| \le a} |\alpha_n(t) - \alpha_n(s)|$$

is the oscillation modulus of α_n.

The proof of (0.3) needs some technical calculations involving both the path and distributional structure of α_n. A survey of available methods may be found in the article of Gaenssler and Stute (1979). The following classification may serve as a brief reference.

I. *The direct method.* In his textbook Billingsley (1968) obtains (0.3) from some general fluctuation inequalities, while in Takács (1967) the result follows from a theorem on processes with interchangeable increments and step functions as sample paths. See also Parthasarathy (1967).

II. *The indirect method.* The problem is one of finding, for each $n \in \mathbb{N}$, a Brownian Bridge B_n° such that with high probability, α_n is close to B_n°. The study of such "strong approximations" was initiated by Brillinger (1969) and Breiman (1968), and by Pyke and Root (1968). While their approach is based on the classical Skorokhod embedding scheme the approximation of Komlós, Major and Tusnády (1975) is obtained from a quantile transformation of appropriate binomial random variables. The error term is determined by the degree of approximation between binomial and normal tails. We should also mention the very insightful proof indicated in Breiman's book, where (0.3) is related to the oscillation of the partial sum process of independent exponential random variables. While I provides a direct proof of (0.3), one can see that in II an estimate for the approximating process corresponding to (0.3) is needed. In summary, this shows that the local behavior of one of these processes is responsible for (b).

This was our primary motivation for the present paper. As a main result we shall show that, in a sense, the uniform empirical process has asymptotically the same local behavior as its limit, the Brownian Bridge.

THEOREM 0.1. *Let $(a_n)_n$ be any sequence in* (0, 1) *with $a_n \downarrow 0$ satisfying*

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

Then

$$\lim_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{2(t-u)\ln a_n^{-1}}} = 1 \qquad \mathbb{P} - \text{a.s.},$$

where $0 < \underline{c} \le \bar{c} < \infty$ *are two preassigned constants.*

The condition $\underline{c}a_n \le t - u$ prevents $t - u$ from being too small. If one admits arbitrarily small intervals the factor $t - u$ has to be replaced by a_n.

THEOREM 0.2. *Under the conditions of Theorem* 0.1 *one has*

$$\lim_{n\to\infty} \omega_n(a_n)/\sqrt{2a_n \ln a_n^{-1}} = 1 \qquad \mathbb{P} - \text{a.s.}$$

Compare Theorem 0.2 with Hölder's condition of Brownian Motion. In fact, (i)–(iii) from above will be needed to guarantee that $\omega_n(a_n)$ has a "normal" behavior. Condition (ii) is always satisfied if $a_n \to 0$ sufficiently slow. On the other hand, (iii) prevents a_n from being too large. As an example, a_n may be equal to $n^{-\lambda}$ times a logarithmic factor ($0 < \lambda < 1$). Since the set of conditions (i)–(iii) will be needed throughout the paper, to simplify somewhat the refereeing, any sequence satisfying the assumptions of Theorem 0.1 will be henceforth called a sequence of bands (bandsequence).

Work is in progress to handle also the case when α_n is "locally Poisson." Theorems 0.1 and 0.2 will be proved, among others, in Section 2. Section 1 contains two well known facts about the Markovian structure of empirical distribution functions and exponential bounds for binomial tails, which will serve to bound the tails of $\omega_n(a)$ (Lemma 2.4). This bound will be essential to show the "lim sup" part in Theorem 0.1 and 0.2, respectively. As one further application we give a simple proof of tightness for the weighted empirical process as considered by O'Reilly (1974). We also included a version of Theorem 0.1 for the non-transformed empirical process $\beta_n(t) := n^{1/2}(F_n(t) - F(t))$, $t \in \mathbb{R}$, pertaining to a sample with d.f.F. This is useful to prove exact rates of convergence for various nonparametric density estimators (Section 4). Finally, Theorem 0.2 may be applied to get exact estimates in the Bahadur representation of the quantile process obtained by Kiefer (Section 3).

1. Two Auxiliary Lemmas. In this section we summarize two well known facts about empirical distribution functions and binomial tails, which will be needed later on. In the following F_n is the empirical d.f. of a sample with uniform distribution on [0, 1].

LEMMA 1.1. *The process* $nF_n(t)$, $0 \le t \le 1$, *is a Markov-process with the following property: under the condition* $nF_n(z) = s$, *the process* $nF_n(t)$, $z \le t \le 1$, *has the same distribution as* $(n - s)F_{n-s}((t - z)/(1 - z)) + s$, $z \le t \le 1$.

In other words given that s points of the sample are less than or equal to z, the process nF_n on $z \le t \le 1$ is stochastically the same (up to the summand s) as that process resulting from distributing $n - s$ points according to the uniform distribution on $z \le t \le 1$.

Let $T_m\uparrow$ be a sequence of finite subsets of [0, 1) such that $T := \bigcup_m T_m$ is dense in [0, 1). By the preceding lemma, nF_n when restricted to T_m is a discrete-time parameter Markov-process for which Strong Markov holds. For the process on [0, 1), we shall get appropriate estimates by letting m tend to infinity and then use the fact that F_n is uniquely determined by its values on T. Finally, we need a sharp upper bound for binomial tails. In this form (1.1) below is also a special case of a general inequality of Bennett (1962).

LEMMA 1.2. *Let* η *have a binomial distribution with parameters* $0 < p < 1$ *and* $n \in \mathbb{N}$. *Then for each* $z > 0$

$$\mathbb{P}(|\eta - np| \ge z) \le 2\exp\{-np[(1 + z/np)\ln(1 + z/np) - z/np\}. \tag{1.1}$$

Substituting $y = z/np$, we get for the exponent in (1.1)

$$-np[(1+y)\ln(1+y) - y] =: H(y).$$

Since

$$H(y) = -np \int_0^y \ln(1+x)\,dx, \qquad y > 0,$$

and

$$\frac{\ln(1+x)}{x} \to 1 \qquad \text{as} \quad x \downarrow 0,$$

we obtain that for each $0 < \delta < 1$ there is some $x_\delta > 0$ such that

$$\mathbb{P}(|\eta - np| \geq z) \leq 2 \begin{cases} \exp[-(1-\delta)z^2/(2np)] & \text{if} \quad z \leq npx_\delta \\ \exp[-(1-\delta)x_\delta z/2] & \text{if} \quad z > npx_\delta. \end{cases} \tag{1.2}$$

2. The Oscillating Behavior of Empirical Processes. In this section, let h be any nonnegative continuous function on the unit interval with $h(t) \neq 0$ for all $0 < t < 1$. In many statistical applications it is necessary to study the weighted empirical process $\alpha_n(t)/h(t)$, $0 < t < 1$. Clearly, if $1/h$ is bounded on [0, 1], then by Donsker's theorem

$$\alpha_n/h \to_{\mathscr{L}} B^\circ/h \qquad \text{as} \quad n \to \infty.$$

For arbitrary h's we obtain the same result when restricting the processes to interior sets of (0, 1). Hence it suffices to study α_n/h at the endpoints of [0, 1] in the case

$$\liminf_{t \to 0} h(t) = 0 = \liminf_{t \to 1} h(t).$$

For symmetry reasons we shall only discuss the critical point $t = 0$. Suppose that h is nondecreasing in a neighborhood of zero with $h(t) \downarrow 0$ as $t \to 0$. It is intuitively clear that the stochastic behavior near zero will depend on how rapidly $h(t)$ converges to zero, as $t \to 0$. If the rate is slowly enough one might guess that the limit is again B°/h. On the other hand, if $h(t)$ is too small, a departure from this limit would only be natural. In the following we may assume w.l.o.g. that h is nondecreasing on the unit interval.

Let $B(\cdot, p; N)$ denote the distribution function of the binomial distribution with parameters $0 < p < 1$ and $N \in \mathbb{N}$, i.e.

$$B(t, p; N) = \sum_{k=0}^{\langle t \rangle} \binom{N}{k} p^k (1-p)^{N-k}.$$

Put

$$B^*(t, p; N) := B(t + Np, p; N).$$

Hence B^* is equal to B centered at expectation. Fix $0 < a < 1$ and $0 \leq q < 1$, and let T be any finite subset of the interval $[qa, a)$. In the following lemma we relate the stochastic behavior of $F_n(\cdot)$ on T to that of $F_n(a)$.

LEMMA 2.1. *In addition to a, q and T, suppose that* $0 < \delta < 1$ *and* $0 < w$ *are such that*

(i) $h(a) - h(qa) \leq \delta h(a)/4$
(ii) $a \leq \delta/4$
(iii) $8a \leq nw^2h^2(a)\delta^2$.

Then

$$\mathbb{P}([F_n(a) - a]/h(a) > w(1-\delta) \mid \sup_{t \in T}[F_n(t) - t]/h(t) > w) \geq 1/2.$$

PROOF. Let $Z = z$ be the smallest point in T, if any, with

$$\frac{F_n(z) - z}{h(z)} = s > w.$$

Write

$$\tilde{S} = n[sh(Z) + Z] = \text{an integer},$$

so that

$$nF_n(z) = \tilde{s}.$$

Use Lemma 1.1 and apply Strong Markov to the finite-time parameter Markov-process $nF_n(t)$, $t \in T$, to get

$$\begin{aligned}
&\mathbb{P}([F_n(a) - a]/h(a) > w(1-\delta) \mid Z = z, nF_n(z) = \tilde{s})\\
&\quad = 1 - B\left(n[a + w(1-\delta)h(a) - sh(z) - z], \frac{a-z}{1-z}; n - \tilde{s}\right)\\
&\quad = 1 - B^*\left(n\left[a + w(1-\delta)h(a) - sh(z) - z - (1 - sh(z) - z)\frac{a-z}{1-z}\right], \frac{a-z}{1-z}; n - \tilde{s}\right)\\
&\quad = 1 - B^*(n[-w\delta h(a) - w(h(z) - h(a)) + H(s)], p; N) = A,
\end{aligned}$$

where $p = (a-z)/(1-z)$, $N = n - \tilde{s}$ and $H(s) = ((a-z)sh(z))/(1-z) + (w-s)h(z)$. Since H is nonincreasing and h nondecreasing we get

$$\begin{aligned}
A &\geq 1 - B^*(n[-w\delta h(a) - w(h(qa) - h(a)) + H(w)], p; N)\\
&\geq 1 - B^*(n[-w\delta h(a) - w(h(qa) - h(a)) + awh(a)], p; N).
\end{aligned}$$

By (i), (ii), and the monotonicity of B^*, the last term is greater than or equal to

$$1 - B^*\left(-nwh(a)\left[\delta - \frac{\delta}{4} - \frac{\delta}{4}\right], p; N\right) \geq 1 - \frac{4Np}{[nwh(a)\delta]^2} \geq 1 - \frac{4a}{n[wh(a)\delta]^2} \geq \frac{1}{2}$$

where the first relation follows from Tschebychev's inequality, and the last is a consequence of (iii). Integrating proves Lemma 2.1. □

The following lemma may be proved by the same arguments as in Lemma 2.1.

LEMMA 2.2. *Under the assumptions of Lemma* 2.1 *we have*

$$\mathbb{P}([F_n(a) - a]/h(a) < -w(1-\delta) \mid \inf_{t\in T}[F_n(t) - t]/h(t) < -w) \geq ½.$$

Lemma 2.1 implies that $\mathbb{P}(\sup_{t\in T}[F_n(t) - t]/h(t) > w) \leq 2\mathbb{P}([F_n(a) - a]/h(a) > w(1-\delta))$, and similarly by Lemma 2.2 $\mathbb{P}(\inf_{t\in T}[F_n(t) - t]/h(t) < -w) \leq 2\mathbb{P}([F_n(a) - a]/h(a) < -w(1-\delta))$. Summation yields $\mathbb{P}(\sup_{t\in T}|F_n(t) - t|/h(t) > w) \leq 2\mathbb{P}(|F_n(a) - a|/h(a) > w(1-\delta))$. It is notable that the right-hand side does not depend on T. Hence if $(T_m)_m$ is any sequence of finite subsets in $[qa, a)$, the last inequality may be applied for each m. In particular, if $T_m \uparrow T$ and T is dense in $[qa, a]$ we obtain, since F_n has sample paths in $D[0, 1]$ and h is continuous,

$$\mathbb{P}(\sup_{qa\leq t\leq a}|F_n(t) - t|/h(t) > w) \leq 2\mathbb{P}(|F_n(a) - a|/h(a) > w(1-\delta)). \tag{2.1}$$

For $h \equiv 1$ condition (i) in Lemma 2.1 is automatically satisfied with $q = 0$. If s is any positive number and if we let $w = s\sqrt{a/n}$, then (iii) is equivalent to the condition $8 \leq s^2\delta^2$. Furthermore, if $s \leq x_\delta\sqrt{na}$, we may apply the first line in (1.2) to bound the right-hand side in (2.1). In summary, we have proved

LEMMA 2.3. *Suppose that for* $0 < a, \delta < 1$ *and* $s > 0$ *we have*

(ii) $a < \delta/4$ (iii) $8 \le (s\delta)^2$ (iv) $s \le x_\delta\sqrt{na}$.

Then

$$\mathbb{P}(\sup_{0\le t\le a} |\alpha_n(t)| > s\sqrt{a}) \le 2\mathbb{P}(|\alpha_n(a)| > s(1-\delta)\sqrt{a}) \le 4\exp[-s^2(1-\delta)^3/2].$$

Recall that the oscillation modulus of α_n is defined by

$$\omega_n(a) = \sup_{|t-s|\le a} |\alpha_n(t) - \alpha_n(s)|.$$

The following lemma provides a sharp exponential upper bound for the tail of $\omega_n(a)/\sqrt{a}$. For the proof we shall use Lemma 2.3 and the fact that α_n has stationary increments.

LEMMA 2.4. *For given* $0 < a, \delta < 1$ *and* $s > 0$ *suppose that*

(ii) $a < \delta/4$ (iii) $8 \le [s\delta/(1+\delta)]^2$ (iv) $s \le \delta x_\delta\sqrt{na}/4$,

then

$$\mathbb{P}(\omega_n(a) > s\sqrt{a}) \le C_\delta a^{-1}\exp[-s^2(1-\delta)^5/2],$$

where $C_\delta = 64\delta^{-2}$ *depends only on* δ.

PROOF. Let R be the smallest positive integer satisfying

$$\delta\sqrt{a}/2 \ge 1/\sqrt{R}.$$

Check that

$$\omega_n(a) \le \max_{0\le i\le R-1}\sup_{0\le t\le a}\left|\alpha_n\left(\frac{i}{R}+t\right) - \alpha_n\left(\frac{i}{R}\right)\right|$$
$$+ 2\max_{0\le i\le R-1}\sup_{0\le\tau<1/R}\left|\alpha_n\left(\frac{i}{R}+\tau\right) - \alpha_n\left(\frac{i}{R}\right)\right|$$

Apply Lemma 2.3 and use stationarity to get

$$\begin{aligned}
&\mathbb{P}(\omega_n(a) > s\sqrt{a})\\
&\le \mathbb{P}\left(\max_{0\le i\le R-1}\sup_{0\le t\le a}\left|\alpha_n\left(\frac{i}{R}+t\right) - \alpha_n\left(\frac{i}{R}\right)\right| > s\sqrt{a}/(1+\delta)\right)\\
&+ \mathbb{P}\left(2\max_{0\le i\le R-1}\sup_{0\le\tau\le 1/R}\left|\alpha_n\left(\frac{i}{R}+\tau\right) - \alpha_n\left(\frac{i}{R}\right)\right| > \delta s\sqrt{a}/(1+\delta)\right)\\
&\le 4R\exp[-s^2(1-\delta)^3/2(1+\delta)^2] + R\mathbb{P}(\sup_{0\le t\le 1/R}|\alpha_n(t)| > s/(1+\delta)\sqrt{R}).
\end{aligned}$$

Since $1/R \le a < \delta/4$ and $s \le \delta x_\delta\sqrt{na}/4 \le x_\delta\sqrt{n}/2\sqrt{R-1} \le x_\delta\sqrt{n/R}$, Lemma 2.3 may be applied to bound the last term in the above estimate. Hence

$$\begin{aligned}
\mathbb{P}(\omega_n(a) > s\sqrt{a}) &\le 8R\exp[-s^2(1-\delta)^3/2(1+\delta)^2]\\
&\le 64\delta^{-2}a^{-1}\exp[-s^2(1-\delta)^5/2]. \quad \square
\end{aligned}$$

As a first application we give a simple proof for (0.3). To this end take any positive ε and η and let $0 < \delta < 1$ be arbitrary, say $\delta = ½$. Let $a > 0$ be chosen so small that

(ii) $a < 1/8$ (iii) $8 \le \frac{1}{9}\sqrt{a}$

(v) $C_{1/2}a^{-1}\exp(-a^{-1/2}/64] \le \eta$ (vi) $\varepsilon \ge a^{1/4}$.

Since for all large enough n

(iv) $a^{-1/4} \le \frac{1}{2}x_{1/2}\sqrt{na}/4$, we may apply Lemma 2.4 with $s = a^{-1/4}$ to get

$$\mathbb{P}(\omega_n(a) > \varepsilon) \le \mathbb{P}(\omega_n(a) > s\sqrt{a}) \le \eta.$$

This proves (0.3) and hence (0.2), in view of the convergence of the finite-dimensional distributions. However, as indicated earlier it is sometimes necessary (see e.g. Pyke and Shorack (1968)) to study the asymptotic behavior of the weighted empirical process α_n/h.

In an equivalent form this amounts to the question, whether α_n converges to $B°$ w.r.t. the stronger metric

$$d_h(f, g) = d(f/h, g/h), \qquad f, g \in D[0, 1],$$

where d is the Skorokhod metric (or sometimes the sup-norm metric) on $D[0, 1]$ (cf. Billingsley (1968)), and h is a preassigned weighting function. As before we assume that h is continuous and positive on (0, 1). Furthermore, let h be nondecreasing (nonincreasing) in some neighborhood of zero (one). Suppose that $h(t) \downarrow 0$ as $t \downarrow 0$ and (or) $h(t) \downarrow 0$ as $t \uparrow 1$, for in the other case everything follows from (0.2). It can be easily seen (using almost surely convergent versions of α_n) that $\alpha_n \to_{\mathscr{L}} B°$ w.r.t. d_h if and only if

(2.2) For each $\varepsilon > 0$ and every $\eta > 0$ there is some $a > 0$ such that for all large $n \in \mathbb{N}$

$$\mathbb{P}(\sup_{0<t<a} |\alpha_n(t)|/h(t) > \varepsilon) \le \eta$$

and

$$\mathbb{P}(\sup_{1-a<t<1} |\alpha_n(t)|/h(t) > \varepsilon) \le \eta.$$

Cibisov (1964) showed (2.2) for the class of h's which are regularly growing at the endpoints of [0, 1] and for which

$$\int_0^1 t^{-1} \exp[-\varepsilon h_i^2(t)/t]\, dt < \infty \qquad \text{for all} \quad \varepsilon > 0, \quad i = 1, 2 \tag{2.3}$$

where $h_1(t) = h(t)$ and $h_2(t) = h(1 - t)$.

O'Reilly (1974) proved (without the regularity assumptions) that (2.3) is both necessary and sufficient for (2.2). While the necessity part follows from a characterization of the lower and upper classes of $B°$, the sufficiency part presents some difficulties. In the above mentioned papers the proof proceeds by associating to α_n an appropriately standardized Poisson process for which a condition like (2.2) follows from (2.3).

Let us show the usefulness of our analysis by giving a proof of (2.2) for which poissonization is superfluous. For symmetry reasons we shall only consider the critical point $t = 0$. To simplify somewhat the arguments we assume that the function $t \to h(t)/\sqrt{t}$ is nonincreasing in a neighborhood of zero.

Now, for given positive ε and η choose $a > 0$ so small that for an arbitrary $0 < \delta < 1$, say $\delta = ½$, the following holds:

for some constant $C > 0$ to be specified later

$$C \int_0^{2a} t^{-1} \exp[-\varepsilon^2(1 - \delta)^3 h^2(t)/2t]\, dt \le \eta/2.$$

This is always possible because of (2.3). Put $w = \varepsilon/\sqrt{n}$. Furthermore, let a be such that

(ii) $a < \delta/4$ and (iii) $8a \le [\varepsilon h(a)\delta]^2$.

Note that by (2.3)

$$\lim_{t\to 0} h^2(t)/t = \infty,$$

so that (iii) is always satisfied for all small enough a. Let $0 < q < 1$ be such that $1 - \sqrt{q} \le \delta/4$. By the monotonicity of $t \to h(t)/\sqrt{t}$

$$1 - \frac{h(q^{m+1}a)}{h(q^m a)} = 1 - \frac{\sqrt{q^{m+1}a}\,h(q^{m+1}a)\sqrt{q^m a}}{\sqrt{q^m a}\sqrt{q^{m+1}a}\,h(q^m a)} \le 1 - \sqrt{q} \le \frac{\delta}{4},$$

that is, (i) holds for all $q^m a$, $m \ge 0$. Furthermore, conditions (ii) and (iii) are also satisfied with a replaced by $q^m a$, $m \ge 0$. Hence (2.1) is applicable for all $q^m a$, $m = 0, 1, \cdots$. Fix n. Let $m_0 \in \mathbb{N}$ be defined by $q^{m_0+1}a \le \rho/n < q^{m_0}a$, where $0 < \rho < 1$ is a constant so that $1 - \exp(-\rho) \le \eta/8$. Now,

$$\mathbb{P}(\sup_{0<t<a} |\alpha_n(t)|/h(t) > \varepsilon) \le \mathbb{P}(\sup_{\rho/n \le t < a} |\alpha_n(t)|/h(t) > \varepsilon)$$
$$+ \mathbb{P}(\sup_{0<t\le\rho/n} |\alpha_n(t)|/h(t) > \varepsilon \equiv A + B.$$

Use (1.2) and (2.1) to bound A by

$$A \le \sum_{m=0}^{m_0} \mathbb{P}(\sup_{q^{m+1}a \le t \le q^m a} |\alpha_n(t)|/h(t) > \varepsilon) \le 2\sum_{m=0}^{m_0} \mathbb{P}(|\alpha_n(q^m a)|/h(q^m a) > \varepsilon(1-\delta))$$

$$\le 4\sum_{m=0}^{m_0}\left[\exp\left(-\frac{\varepsilon^2(1-\delta)^3 h^2(q^m a)}{2q^m a}\right) + \exp(-c_0\sqrt{n}h(q^m a)\varepsilon)\right] \qquad (c_0 := (1-\delta)^2 x_\delta/2)$$

$$\le \frac{4}{1-q}\int_0^{2a} t^{-1}\exp\left[-\frac{\varepsilon^2(1-\delta)^3 h^2(t)}{2t}\right]dt + 4\sum_{m=0}^{m_0}\exp(-c_0\varepsilon h(a)\sqrt{\rho q^{m-m_0}/a})$$

$$\le \eta/2 + 4\sum_{i=0}^{\infty}\exp -c_0\varepsilon h(a)\sqrt{\rho/a}q^{-i/2}),$$

if we put $C = 4/(1-q)$. Since the last series converges for each a and $h(a)/\sqrt{a} \to \infty$ as $a \to 0$, the last term can be made less than $\eta/4$ upon choosing a sufficiently small. Hence $A \le 3\eta/4$. It suffices to show

$$B \le \eta/4 \qquad \text{for all large } n \in \mathbb{N}. \tag{2.4}$$

Let $\xi_{1:n} := \min(\xi_1, \cdots, \xi_n)$. Then

$$\mathbb{P}(\xi_{1:n} \le \rho/n) = 1 - \mathbb{P}(\xi_{1:n} > \rho/n) = 1 - (1-\rho/n)^n \to 1 - \exp(-\rho) \le \eta/8,$$

whence

$$\mathbb{P}(\xi_{1:n} \le \rho/\mathrm{n}) \le \eta/4 \qquad \text{for all } n \ge n_0(\rho), \text{ say.}$$

On the set $\{\xi_{1:n} > \rho/n\}$ $\sup_{0<t\le\rho/n}|\alpha_n(t)|/h(t) = \sqrt{n}\sup_{0<t\le\rho/n} t/h(t) \le \sqrt{\rho}\sup_{0<t\le\rho/n}\sqrt{t}/h(t) \to 0$, as $n \to \infty$. This shows (2.4) and completes the proof of (2.2).

Up to a logarithmic term $h_0(t) = [t(1-t)]^{1/2}$ is the function for which (2.3) just fails to hold. This is in accordance with a recent result obtained by Jaeschke (1979) who showed that for some constants γ_n and ρ_n the variable

$$\gamma_n \sup_{0<t<1}|\alpha_n(t)|/h_0(t) - \rho_n$$

has in the limit an extreme value distribution. If h is even smaller than h_0 it will be necessary to restrict the supremum on subintervals $(\varepsilon_n, 1-\varepsilon_n)$, where $\varepsilon_n \downarrow 0$ sufficiently slowly. We shall not discuss this in greater detail since it is beyond of the scope of the present paper.

For investigating the local behavior of the empirical process the following estimate will be useful. Shorack and Wellner (1982) adopted our technique to prove analogous results also for more general weighting functions.

LEMMA 2.5. In the notation of Lemma 2.4 (with $0 < q < 1$), suppose that for $0 < \underline{c} \le \bar{c} < \infty$

(ii) $\bar{c}a < \delta/4$ (iii) $8 \le s^2 q\delta^2/(1+\delta)^2$ (iv) $s\sqrt{q} \le \delta x_\delta\sqrt{n\underline{c}a}/4$.

Then for some constant $C > 0$ depending only on $\underline{c}$, $\bar{c}$, q and δ

$$\mathbb{P}(\sup_{\underline{c}a \le t-u \le \bar{c}a}|\alpha_n(t) - \alpha_n(u)|/\sqrt{t-u} > s) \le Ca^{-1}\exp[-s^2 q(1-\delta)^5/2]. \tag{2.5}$$

PROOF. Let $m_0 \in \mathbb{N} \cup \{0\}$ be defined by the relation $q^{m_0+1}\bar{c} \le \underline{c} \le q^{m_0}\bar{c}$. Apply Lemma 2.4 to get

$$\begin{aligned} &\mathbb{P}(\sup_{\underline{c}a\le t-u\le\bar{c}a}|\alpha_n(t)-\alpha_n(u)|/\sqrt{t-u}>s) \\ &\quad\le \textstyle\sum_{m=0}^{m_0}\mathbb{P}(\sup_{q^{m+1}\bar{c}a\le t-u\le q^m\bar{c}a}|\alpha_n(t)-\alpha_n(u)|/\sqrt{t-u}>s) \\ &\quad\le \textstyle\sum_{m=0}^{m_0}\mathbb{P}(\omega_n(q^m\bar{c}a)>s\sqrt{q}\sqrt{q^m\bar{c}a}) \\ &\quad\le \textstyle\sum_{m=0}^{m_0}C_\delta(q^m\bar{c}a)^{-1}\exp[-s^2q(1-\delta)^5/2], \end{aligned}$$

whence the assertion. □

In the case $\underline{c} = \bar{c}$ (2.5) is also true for $q = 1$.

By Lemma 2.5 it is now possible to derive $\mathbb{P}$-a.s. asymptotic upper bounds for the oscillation of α_n. Let $0 < \underline{c} \le \bar{c} < \infty$ be fixed throughout.

LEMMA 2.6. Let $0 < a_n < 1$ be any bandsequence as in Theorem 0.1, i.e. $a_n \downarrow 0$ and
(i) $na_n \uparrow \infty$ (ii) $\ln a_n^{-1} = o(na_n)$ (iii) $\ln a_n^{-1}/\ln\ln n \to \infty$.
Then with probability one

$$\limsup_{n\to\infty}\sup_{\underline{c}a_n\le t-u\le\bar{c}a_n}\frac{|\alpha_n(t)-\alpha_n(u)|}{\sqrt{2(t-u)\ln a_n^{-1}}}\le 1. \tag{2.6}$$

PROOF. For given $\varepsilon > 0$ let $s_n := \sqrt{2(1+\varepsilon)\ln a_n^{-1}}$. Define

$$A_n := \left\{\sup_{\underline{c}a_n\le t-u\le\bar{c}a_n}\frac{|\alpha_n(t)-\alpha_n(u)|}{\sqrt{t-u}}>s_n\right\}.$$

For the proof of (2.6) it remains to show

$$\mathbb{P}(\limsup_{n\to\infty}A_n) = 0. \tag{2.7}$$

For this, let $0 < \rho < 1$ be a constant to be specified later, and put $n_k := \langle(1+\rho)^k\rangle$, $k \in \mathbb{N} \cup \{0\}$. Clearly as $k \to \infty$,

$$n_k/n_{k-1} \to 1+\rho \qquad\text{and}\qquad \ln n_k/\ln n_{k-1} \to 1.$$

Finally, for each k and every $n_{k-1} \le n < n_k$ we set

$$C_k := \left\{\sup_{\underline{c}a_{n_k}\le t-u\le 2\bar{c}(1+\rho)a_{n_k}}\frac{|\alpha_{n_{k+1}}(t)-\alpha_{n_{k+1}}(u)|}{\sqrt{t-u}}>\left(1+\frac{\varepsilon}{8}\right)\sqrt{2\ln a_{n_k}^{-1}}\right\}$$

and

$$B_{n,n_{k+1}} := \{\omega_{n,n_{k+1}}(\bar{c}a_n) \le 2\sqrt{\bar{c}a_n\ln(\bar{c}a_n)^{-1}}\},$$

where $\omega_{n,n_{k+1}}$ is the oscillation modulus for the sample $\xi_{n+1}, \cdots, \xi_{n_{k+1}}$. It follows from (ii) that uniformly in $n_{k-1} \le n < n_k$

$$\ln a_n^{-1} = o((n_{k+1}-n)a_n) \qquad\text{as } k\to\infty.$$

Hence we may apply Lemma 2.4 for all large k to obtain a lower bound for the probability of $B_{n,n_{k+1}}$. In particular, if $0 < \delta$ is chosen so small that $2(1-\delta)^5 \ge 3/2$, we get $\mathbb{P}(B_{n,n_{k+1}}) \ge 1 - C_\delta(\bar{c}a_n)^{1/2}$ and therefore $\mathbb{P}(B_{n,n_{k+1}}) \ge 1/2$ for $n \ge n_0$, say.

Let us show that for all large k

$$\textstyle\bigcup_{n=n_{k-1}}^{n_k-1}A_n\cap B_{n,n_{k+1}} \subset C_k. \tag{2.8}$$

For this note that for each pair (t, u) with $\underline{c}a_n \le t-u \le \bar{c}a_n$ one has from (i) $\underline{c}a_{n_k} \le t-u \le 2\bar{c}(1+\rho)\,a_{n_k}$ [k large]. Furthermore, by additivity of nF_n,

$$\begin{aligned} n_{k+1}(F_{n_{k+1}}(t)-F_{n_{k+1}}(u)-(t-u)) &= n(F_n(t)-F_n(u)-(t-u)) \\ &\quad+(n_{k+1}-n)(F_{n,n_{k+1}}(t)-F_{n,n_{k+1}}(u)-(t-u)), \end{aligned}$$

where, according to the definition of $\omega_{n,n_{k+1}}$, $F_{n,n_{k+1}}$ is the empirical distribution function for the sample $\xi_{n+1}, \cdots, \xi_{n_{k+1}}$. As $n_{k+1}/n_k \to 1 + \rho$ for $k \to \infty$, we therefore obtain on $A_n \cap B_{n,n_{k+1}}$ [k large]:

$$\sup_{\underline{c}a_{n_k} \le t-u \le 2\bar{c}(1+\rho)a_{n_k}} \frac{|\alpha_{n_{k+1}}(t) - \alpha_{n_{k+1}}(u)|}{\sqrt{t-u}} \ge \left(\frac{n}{n_{k+1}}\right)^{1/2} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{t-u}}$$

$$-\left(\frac{n_{k+1} - n}{n_{k+1}\underline{c}a_n}\right)^{1/2} \omega_{n,n_{k+1}}(\bar{c}a_n)$$

$$\ge (1 + 2\rho)^{-1}s_n - 2\left[\frac{(n_{k+1} - n)\bar{c}a_n \ln(\bar{c}a_n)^{-1}}{n_{k+1}\underline{c}a_n}\right]^{1/2}$$

$$\le (1 + 2\rho)^{-1}s_n - 2[4\rho\bar{c}\ln(\bar{c}a_n)^{-1}/\underline{c}]^{1/2} =: k_n.$$

Suppose that $\rho > 0$ has been chosen so small that

$$\frac{\sqrt{1+\varepsilon}}{1 + 2\rho} - 2[2\rho\bar{c}/\underline{c}]^{1/2} > 1 + \varepsilon/4,$$

then if n is large enough

$$k_n \ge (1 + \varepsilon/4)\sqrt{2\ln a_n^{-1}} \ge (1 + \varepsilon/4)\sqrt{2\ln[n_{k-1}/n_k a_{n_k}]} \sim (1 + \varepsilon/4)\sqrt{2\ln a_{n_k}^{-1}}.$$

In summary, this proves (2.8). We shall use this and the fact that $A_{n_{k-1}}, \cdots, A_n$ are independent of $B_{n,n_{k+1}}$ to prove the following inequality:

$$\mathbb{P}(\cup_{n=n_{k-1}}^{n_k-1} A_n) \le 2\mathbb{P}(C_k) \qquad [k \text{ large}].$$

In fact, since $\mathbb{P}(B_{n,n_{k+1}}) \ge 1/2$ for sufficiently large k, by (2.8),

$$\mathbb{P}(G_k) \ge \mathbb{P}(\cup_{n=n_{k-1}}^{n_k-1}(A_n \cap B_{n,n_{k+1}})) = \sum_{n=n_{k-1}}^{n_k-1} \mathbb{P}((A_n \cap B_{n,n_{k+1}})\backslash(\cup_{m=n_{k-1}}^{n-1} A_m \cap B_{m,n_{k+1}}))$$

$$\ge \sum_{n=n_{k-1}}^{n_k-1} \mathbb{P}((A_n \cap B_{n,n_{k-1}})\backslash\cup_{m=n_{k-1}}^{n-1} A_m) \ge 1/2\, \mathbb{P}(\cup_{n=n_{k-1}}^{n_k-1} A_n),$$

whence the assertion. By Borel-Cantelli, the proof of (2.7) will be finished by showing $\sum_k \mathbb{P}(C_k) < \infty$. However, this is an easy consequence of Lemma 2.5, if one chooses $0 < \delta$ so small and $0 < q < 1$ so large that

$$q(1 - \delta)^5(1 + \varepsilon/8)^2 \ge 1 + \varepsilon/16.$$

In this case

$$\mathbb{P}(C_k) \le Ca_{n_k}^{-1}\exp[-(1 + \varepsilon/16)\ln a_{n_k}^{-1}] = Ca_{n_k}^{\varepsilon/16}.$$

Since $(n_k)_k$ is geometrically increasing (iii) implies $\sum_k a_{n_k}^{\varepsilon/16} < \infty$. This completes the proof of Lemma 2.6. □

To prove Theorem 0.1 it remains to show that the "lim sup" in (2.6) is in fact a "lim". For this we need two lemmas of which the first is well known. Namely, that $n^{1/2}\alpha_n$ is a centered Poisson process with parameter n, under the condition that n events have been observed up to time 1.

LEMMA 2.7. *Let $N_n(t)$, $0 \le t \le 1$, be a Poisson process with parameter n, and let $N_n'(t) := N_n(t) - nt$. Then*

$$\mathbb{P}(n^{1/2}\alpha_n \in \cdot) = \mathbb{P}(N_n' \in \cdot \mid N_n(1) = n).$$

In using the above representation of the empirical process, it will be necessary to have upper bounds for the Poisson distribution function. To this end the following lemma will be useful as it relates Poisson tails to their corresponding normal ones.

LEMMA 2.8. (Bohman (1963)). *Suppose that* η *has a Poisson distribution with parameter* $\lambda > 0$. *Then for each* $k \in \mathbb{R}$

$$\mathbb{P}((\eta - \lambda)/\sqrt{\lambda} \le k) \le \mathbb{P}(\mathcal{N}(0, 1) \le k + \lambda^{-1/2}),$$

where $\mathcal{N}(0, 1)$ *is a standard normal variate with expectation* 0 *and variance* 1.

It is a notable though simple fact that, by the CLT,

$$\mathbb{P}((\eta - \lambda)/\sqrt{\lambda} \le k) \to \mathbb{P}(\mathcal{N}(0, 1) \le k) \qquad \text{as } \lambda \to \infty.$$

LEMMA 2.9. *For each bandsequence* $(a_n)_n$

$$\liminf_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{2(t-u)\ln a_n^{-1}}} \ge 1 \qquad \mathbb{P}\text{-a.s.} \tag{2.9}$$

PROOF. Consider any partition

$$0 = c_1^n < c_2^n < \cdots < c_{m_n+1}^n = 1$$

of the unit interval into m_n subintervals with length $\underline{c}a_n \le c_{i+1}^n - c_i^n \le \bar{c}a_n$ for all $i = 1, \cdots, m_n$. For given $0 < \varepsilon < 1$ define

$$D_n := D_n(\varepsilon) := \left\{ \sup_{i=1,\ldots,m_n} \frac{\alpha_n(c_{i+1}^n) - \alpha_n(c_i^n)}{\sqrt{c_{i+1}^n - c_i^n}} \le \sqrt{(1-\varepsilon)2\ln a_n^{-1}} \right\}.$$

By Lemma 2.7,

$$\mathbb{P}(D_n) = \mathbb{P}\left(\sup_{i=1,\ldots,m_n} \frac{N_n'(c_{i+1}^n) - N_n'(c_i^n)}{\sqrt{c_{i+1}^n - c_i^n}} \le \sqrt{(1-\varepsilon)2n\ln a_n^{-1}} \,|\, N_n(1) = n \right)$$

$$\le \left[\prod_{i=1}^{m_n} \mathbb{P}\left(\frac{N_n'(c_{i+1}^n) - N_n'(c_i^n)}{\sqrt{c_{i+1}^n - c_i^n}} \le \sqrt{(1-\varepsilon)2n\ln a_n^{-1}} \right) \right] \Big/ \mathbb{P}(N_n(1) = n) =: A.$$

For the last inequality use the fact that N_n has independent increments. To estimate the denominator, apply Stirling's formula to get

$$\mathbb{P}(N_n(1) = n) = \frac{n^n}{n!} e^{-n} \sim (2\pi n)^{-1/2}.$$

Hence by Lemma 2.8 (applied with $\lambda = n(c_{i+1}^n - c_i^n)$)

$$A \le \text{const}\cdot n^{1/2} \prod_{i=1}^{m_n} \mathbb{P}\left(\mathcal{N}(0, 1) \le \sqrt{(1-\varepsilon)2\ln a_n^{-1}} + \frac{1}{\sqrt{\underline{c}na_n}} \right)$$

$$= \text{const}\cdot n^{1/2}[\mathbb{P}(\mathcal{N}(0, 1) \le \sqrt{(1-\varepsilon)2\ln\cdot a_n^{-1}} + (\underline{c}na_n)^{-1/2})]^{m_n}.$$

Since $a_n \to 0$ and $na_n \to \infty$, we therefore obtain

$$\mathbb{P}(D_n) = \mathcal{O}(n^{1/2}[\mathbb{P}(\mathcal{N}(0, 1) \le \sqrt{(1-\varepsilon/2)2\ln a_n^{-1}})]^{m_n})$$

$$= \mathcal{O}\left(n^{1/2} \left[1 - \frac{a_n^{1-\varepsilon/2}}{2\pi\sqrt{\ln a_n^{-1}}} \right]^{m_n} \right) = \mathcal{O}(n^{1/2}\exp(-m_n a_n^{1-\varepsilon/4})).$$

Furthermore, $(m_n a_n)_n$ is bounded against zero, so that

$$\mathbb{P}(D_n) = \mathcal{O}(n^{1/2}\exp(-a_n^{-\varepsilon/8})).$$

Condition (iii) of Lemma 2.6 implies that $\sum \mathbb{P}(D_n) < \infty$, i.e. we have shown $\mathbb{P}(\limsup_{n\to\infty} D_n) = 0$. This yields

$$\liminf_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n} \frac{\alpha_n(t) - \alpha_n(u)}{\sqrt{2(t-u)\ln a_n^{-1}}} \ge \sqrt{1-\varepsilon} \qquad \mathbb{P}\text{-a.s.}$$

Letting ε tend to zero proves Lemma 2.9. □

Note that for the last result condition (ii) in Lemma 2.6 and the monotonicity assumptions on $(a_n)_n$ were superfluous. It is also a notable fact that the same arguments hold for arbitrary (non-degenerate) subintervals J of [0, 1], if instead of [0, 1], $c_1^n < \cdots < c_{m_n+1}^n$ is a partition of J. By letting ε tend to zero we therefore get

$$\liminf_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\, t,u\in J} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{2(t-u)\ln a_n^{-1}}} \ge 1 \qquad \mathbb{P}\text{-a.s.} \tag{2.9*}$$

Together with Lemma 2.6 this gives the following extension of Theorem 0.1.

THEOREM 2.10. *For each bandsequence $(a_n)_n$ as in Theorem 0.1 and every subinterval J of [0, 1] one has*

$$\lim_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\, t,u\in J} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{2(t-u)\ln a_n^{-1}}} = 1 \qquad \mathbb{P}\text{-a.s.} \tag{2.10}$$

The interval J may be replaced by any subset of [0, 1] with nonempty interior. For a sample with distribution function F the corresponding result may easily be derived by considering the version (0.1). Since the set J now transforms into the set $F(J)$ we have to assume that $F(J)$ has a nonempty interior. As may be seen from a one point distribution a condition like this is necessary to avoid trivial statements.

Consider the "non-transformed" empirical process β_n pertaining to $\xi_1, \cdots, \xi_n$:

$$\beta_n(t) := n^{1/2}(F_n(t) - F(t)), \qquad t \in \mathbb{R}.$$

THEOREM 2.11. *Suppose that $F(J)$ has a nonempty interior. Then for each bandsequence $(a_n)_n$*

$$\lim_{n\to\infty} \sup_{\underline{c}a_n \le F(t)-F(u) \le \bar{c}a_n;\, t,u\in J} \frac{|\beta_n(t) - \beta_n(u)|}{\sqrt{2[F(t)-F(u)]\ln a_n^{-1}}} = 1 \qquad \mathbb{P}\text{-a.s.} \tag{2.11}$$

In particular, if F is differentiable on J and if $f = F'$ is bounded against 0 and ∞:

$$0 < m \le f(x) \le M < \infty \qquad \text{for all } x \in J,$$

then $\underline{c}a_n \le t - u \le \bar{c}a_n$ implies $m\underline{c}a_n \le F(t) - F(u) \le M\bar{c}a_n$. Hence by (2.11) (with $\underline{c}$ and $\bar{c}$ replaced by $m\underline{c}$ and $M\bar{c}$, respectively) $\mathbb{P}$-almost surely

$$\limsup_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\, t,u\in J} \frac{|\beta_n(t) - \beta_n(u)|}{\sqrt{2[F(t)-F(u)]\ln a_n^{-1}}} \le 1.$$

For the lower part consider a partition $J_1, \cdots, J_{m_n}$ of J into subintervals each of which has length $\underline{c}a_n \le \ell(J_i) \le \bar{c}a_n$. By assumption $F(J_1), \cdots, F(J_{m_n})$ is a partition of $F(J)$ with $m\underline{c}a_n \le \ell(F(J_i)) \le M\bar{c}a_n$. The same arguments which led to (2.9) now yield that for the particular version (0.1) and hence for general β_n

$$\liminf_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\, t,u\in J} \frac{|\beta_n(t) - \beta_n(u)|}{\sqrt{2[F(t)-F(u)]\ln a_n^{-1}}} \ge 1 \qquad \mathbb{P}\text{-a.s.}$$

Thus we have proved

THEOREM 2.12. *Suppose that F is differentiable on J with $0 < m \le F'(x) = f(x) \le M < \infty$ for all $x \in J$. Then for each bandsequence $(a_n)_n$*

$$\lim_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\, t,u\in J} \frac{|\beta_n(t) - \beta_n(u)|}{\sqrt{2[F(t)-F(u)]\ln a_n^{-1}}} = 1 \qquad \mathbb{P}\text{-a.s.}$$

By the mean value theorem

$$F(t) - F(u) = f(y)(t-u) \qquad \text{for some } u \le y \le t.$$

Under mild conditions on f, it is possible to replace y by an arbitrary point $u \le x_{u,t} \le t$. For example, suppose that f is uniformly continuous on J. Since $f \ge m > 0$ on J and $t - u \le \bar{c}a_n \to 0$

$$f(y) = f(x_{u,t})[1 + o(1)] \qquad \text{uniformly in } u \text{ and } t.$$

THEOREM 2.13. *Suppose that on J, F has a uniformly continuous derivative f such that $0 < m \le f(x) \le M < \infty$ for all $x \in J$. Let $x_{u,t}$ be any point between u and t. Then for each bandsequence $(a_n)_n$*

$$\lim_{n\to\infty} \sup_{\underline{c}a_n \le t-u \le \bar{c}a_n;\ t,u\in J} \frac{|\beta_n(t) - \beta_n(u)|}{\sqrt{2(t-u)f(x_{u,t})\ln a_n^{-1}}} = 1 \qquad \mathbb{P}\text{-a.s.}$$

If F has finite support, i.e. when $F(x_0) = 0$ and $F(x_1) = 1$ for some real x_0, x_1 it may happen that the assumptions of Theorem 2.13 hold with $J = [x_0, x_1]$. If F has nonfinite support, the situation is quite different. For example, suppose that F has a uniformly continuous (positive) derivative f on the real line. Then $f(x) \to 0$ as $x \to \pm\infty$. So the boundedness condition $0 < m \le f(x)$ excludes one from setting $J = \mathbb{R}$ (or any unbounded subinterval of $\mathbb{R}$). However, since such distributions play an important role in many fields of statistical application, it would be useful to get related results also in this case. To this end, consider the uniform empirical process α_n and take $\underline{c} = 1 = \bar{c}$. By Theorem 2.10,

$$\lim_{n\to\infty} \sup_{t-u=a_n} \frac{|\alpha_n(t) - \alpha_n(u)|}{\sqrt{2a_n \ln a_n^{-1}}} = 1 \qquad \mathbb{P}\text{-a.s.}$$

whence

$$\liminf_{n\to\infty} \omega_n(a_n)/\sqrt{2a_n \ln a_n^{-1}} \ge 1 \qquad \mathbb{P}\text{-a.s.}$$

To show that the "lim inf" is in fact a "lim" it remains to prove that for $\tilde{A}_n := \{\omega_n(a_n) > \sqrt{2(1+\varepsilon)\, a_n \ln a_n^{-1}}\}$, $\varepsilon > 0$, one has $\mathbb{P}(\limsup_{n\to\infty} \tilde{A}_n) = 0$. However, this follows by using the same method of proof as in Lemma 2.6. Hence we obtain the following result, which is identical with our Theorem 0.2.

THEOREM 2.14. *For the uniform empirical process and bands a_n*

$$\lim_{n\to\infty} \omega_n(a_n)/\sqrt{2a_n \ln a_n^{-1}} = 1 \qquad \mathbb{P}\text{-a.s.} \tag{2.12}$$

If F is differentiable on J, (2.12) implies (use (0.1))

$$\limsup_{n\to\infty} \sup_{|t-u|\le a_n;\ t,u\in J} |\beta_n(t) - \beta_n(u)|/\sqrt{2a_n \ln a_n^{-1}} \le \sqrt{\sup_{x\in J} F'(x)} \qquad \mathbb{P}\text{-a.s.} \tag{2.13}$$

For sufficiently smooth F' (2.13) can be strengthened as follows.

THEOREM 2.15. *Suppose that on J F has a (positive) uniformly continuous derivative f. Then for each bandsequence $(a_n)_n$*

$$\lim_{n\to\infty} \sup_{|t-u|\le a_n;\ t,u\in J} |\beta_n(t) - \beta_n(u)|/\sqrt{2a_n \ln a_n^{-1}} = \sqrt{\sup_{x\in J} f(x)} \qquad \mathbb{P}\text{-a.s.}$$

PROOF. By (2.13) it suffices to prove the "lim inf" -part. Since f is uniformly continuous and has finite integral we get $M := \sup_{x\in J} f(x) < \infty$. Furthermore, we may choose a subinterval J_o of J such that for given $0 < \varepsilon < M/2$

$$f(x) \ge M - \varepsilon > 0 \qquad \text{for all } x \in J_0.$$

For this J_o we may apply Theorem 2.13 to get (with $\underline{c} = 1 = \bar{c}$)

$$\liminf_{n\to\infty} \sup_{|t-u|\le a_n;\, t,u\in J_0} |\beta_n(t) - \beta_n(u)|/\sqrt{2a_n \ln a_n^{-1}} \ge \sqrt{M - \varepsilon} \qquad \mathbb{P}\text{-a.s.}$$

Since $J_0 \subset J$ and $\varepsilon > 0$ was arbitrary this proves the assertion. □

3. Approximation of Empirical Quantiles. For each d.f. F on the real line the corresponding inverse function has been defined by

$$F^{-1}(p) = \inf\{t \in \mathbb{R}: F(t) \ge p\}, \qquad 0 < p < 1.$$

In many procedures it is desirable to have some knowledge of F^{-1} rather than F itself. Intuitively one might think that, as $F_n(t)$ is a good approximation of $F(t)$, the "empirical quantile" $F_n^{-1}(p)$ could as well serve as a statistical substitute for $F^{-1}(p)$. In fact, under some mild smoothness assumptions on F this is always possible uniformly in p within a fixed interior set of (0, 1), i.e. when p is bounded from zero and one. However, if we let p be arbitrarily close to 0 or 1, the deviation between F^{-1} and F_n^{-1} may be too large. This stems from the fact that the asymptotic behavior of $F_n^{-1}(p) - F^{-1}(p)$ depends on the local behavior of F at $F^{-1}(p)$. To make this clear we first consider the case when $F_n^{-1}(p)$, $0 < p < 1$, is the empirical quantile of a uniform sample on [0, 1]. It is easy to check that

$$\sup_{0\le t\le 1} |F_n(t) - t| = \sup_{0<p<1} |F_n^{-1}(p) - p|. \tag{3.1}$$

By the Glivenko-Cantelli Theorem we therefore obtain that the maximal deviation between F_n^{-1} and F^{-1} tends to zero with probability one, as $n \to \infty$. To get more detailed information on the asymptotic behavior of $F_n^{-1}(p)$, Bahadur (1966) initiated to study the effective error $F_n^{(2)}(p) - F_n^{-1}(p)$, $0 < p < 1$, where $F_n^{(2)}(p)$ is defined by

$$F_n^{(2)}(p) = p - (F_n(p) - p), \qquad 0 < p < 1.$$

He showed that for fixed $0 < p < 1$

$$|F_n^{(2)}(p) - F_n^{-1}(p)| = \mathcal{O}(n^{-3/4}(\ln n)^{1/2}(\ln\ln n)^{1/4}) \qquad \mathbb{P}\text{-a.s.}$$

Kiefer (1967) proved that the exact rate is of order $n^{-3/4}(\ln\ln n)^{3/4}$. Hence $n^{1/2}(F_n^{-1}(p) - p)$ has the same asymptotic behavior as $n^{1/2}(p - F_n(p))$ for each fixed p. To obtain functional limit theorems, one has to study the maximal deviation between $F_n^{(2)}$ and F_n^{-1}.

As to this, Kiefer (1970) showed that with probability one

$$\limsup_{n\to\infty} \sup_{0<p<1} n^{3/4}(\ln\ln n)^{-1/4}(\ln n)^{-1/2} |F_n^{(2)}(p) - F_n^{-1}(p)| = 2^{-1/4}. \tag{3.2}$$

From this it follows that the (uniform) quantile process

$$q_n(p) := n^{1/2}(F_n^{-1}(p) - p), \qquad 0 < p < 1,$$

obeys the same invariance principles as the uniform empirical process. Clearly, to derive such a result it suffices to prove any constant in (3.2) with "≤" rather than equality. Let us show that (with $2^{-1/4}$) this is an easy consequence of Theorem 0.2 and the Chung-Smirnov LIL for the empirical d.f. (cf. Gaenssler and Stute (1979), page 202).

Actually, given $\varepsilon > 0$, define $a_n := (1 + \varepsilon)[\ln\ln n/2n]^{1/2}$, $n \ge 3$. Then $(a_n)_n$ is a bandsequence for which by (3.1) and the above mentioned results

$$\limsup_{n\to\infty} \sup_{0<p<1} \frac{|\alpha_n(F_n^{-1}(p)) - \alpha_n(p)|}{\sqrt{2a_n \ln a_n^{-1}}} \le 1 \qquad \mathbb{P}\text{-a.s.}$$

Since

$$\left(\frac{n}{2a_n \ln a_n^{-1}}\right)^{1/2} \sim \frac{2^{1/4} n^{3/4}}{(1+\varepsilon)^{1/2}(\ln\ln n)^{1/4}(\ln n)^{1/2}}$$

and $\varepsilon > 0$ was arbitrary, we get $\mathbb{P}$-a.s.

$$\limsup_{n\to\infty} \sup_{0<p<1} \frac{2^{1/4} n^{3/4} |F_n(F_n^{-1}(p)) - F_n(p) - F_n^{-1}(p) + p|}{(\ln\ln n)^{1/4}(\ln n)^{1/2}} \le 1.$$

The upper bound in (3.2) now follows from the observation that

$$\sup_{0<p<1} | F_n(F_n^{-1}(p)) - p | \le n^{-1}.$$

A close look at Kiefer's (1970) original proof of (3.2) will show that his method is somewhat related to our technique. However, from our asymptotic approach it is more or less straightforward to see why the peculiar norming factor occurs.

By iteration it is possible to get even better estimates for $F_n^{-1}(p)$. As a k-th, $k \ge 3$, approximation we put

$$F_n^{(k)}(p) = F_n^{(k-1)}(p) - F_n(F_n^{(k-1)}(p)) + p,$$

i.e. within $1/n$, we have

$$F_n^{(k)}(p) - F_n^{-1}(p) = n^{-1/2}[\alpha_n(F_n^{-1}(p)) - \alpha_n(F_n^{(k-1)}(p))]$$

The analogue of (3.2) is only stated for the case $k = 3$, but it presents no difficulties to treat a general k:

$$\text{(3.3)} \qquad \limsup_{n\to\infty} \sup_{0<p<1} n^{7/8}(\ln\ln n)^{-1/8}(\ln n)^{-3/4} | F_n^{(3)}(p) - F_n^{-1}(p) | \le \left[\frac{3}{2^{5/4}}\right]^{1/4} \quad \mathbb{P}\text{-a.s.}$$

By the Hewitt-Savage zero-one law, the left-hand side of (3.3) is equal to some constant $C \le (3/2^{5/4})^{1/4}$. However, our method does not provide anything about the true value of C.

Suppose now that the sample comes from an arbitrary d.f. F. The empirical quantile may be easily expressed in terms of F^{-1} and the corresponding uniform empirical quantile, if for F_n we consider the particular version (0.1). We then have

$$\text{(3.4)} \qquad F_n^{-1}(p) = F^{-1}(\bar{F}_n^{-1}(p)), \qquad 0 < p < 1.$$

Assume that F has a continuous first derivative f, which is positive on the support of F. In this general case the empirical quantile process is defined by

$$q_n(p) = n^{1/2} f(F^{-1}(p))(F_n^{-1}(p) - F^{-1}(p)), \qquad 0 < p < 1.$$

Putting $\mathscr{U} \equiv f \circ F^{-1}$, the so-called density-quantile function (cf. Parzen (1979)), we then get from (3.4)

$$\begin{aligned} q_n(p) &= n^{1/2}\mathscr{U}(p)(F^{-1}(\bar{F}_n^{-1}(p)) - F^{-1}(p)) \\ &= \bar{q}_n(p)\frac{\mathscr{U}(p)}{\mathscr{U}(\eta)}, \end{aligned}$$

where $\bar{q}_n$ is the quantile process pertaining to $\bar{F}_n^{-1}$ and η is some point between p and $\bar{F}_n^{-1}(p)$. Write

$$\text{(3.5)} \qquad q_n(p) = \bar{q}_n(p) + \left[\frac{\mathscr{U}(p)}{\mathscr{U}(\eta)} - 1\right]\bar{q}_n(p), \qquad 0 < p < 1.$$

From (3.2) we get $\bar{q}_n \to_{\mathscr{L}} B^\circ$ as $n \to \infty$, where as before B° is a Brownian Bridge and "$\mathscr{L}$" denotes convergence in distribution in the space $D[0, 1]$. Since $\bar{F}_n^{-1}(p)$ converges to p and f is continuous and positive on the support of F, the second summand in (3.5) converges to zero in probability uniformly in p within each fixed compact subinterval of $(0, 1)$. Hence to show $q_n \to_{\mathscr{L}} B^\circ$ it suffices to prove that for given $\varepsilon > 0$ there exists some (small) $a > 0$ such that for all large $n \in \mathbb{N}$

$$\mathbb{P}\left(\sup_{0<p\le a} \left|\frac{\mathscr{U}(p)}{\mathscr{U}(\eta)} - 1\right| |\bar{q}_n(p)| \ge \varepsilon\right) \le \varepsilon,$$

and similarly, for $1 - a \le p < 1$ (cf. (2.2)). Suppose that $\mathscr{U}$ is nondecreasing in some neighborhood of 0. It is known (cf., e.g., Wellner (1978), Remark 1), that for some (large) $\lambda \ge 1$

$$p/\lambda \le \bar{F}_n^{-1}(p)$$

with probability greater than or equal to $1 - \varepsilon/2$. Hence it remains to show that for some small $a > 0$ and all large $n \in \mathbb{N}$ (note that $\{\bar{q}_n\}$ is tight)

$$\mathbb{P}(\sup_{0<p\leq a} \frac{|\bar{q}_n(p)|}{h_\lambda(p)} \geq \varepsilon) \leq \varepsilon/2, \tag{3.6}$$

where $h_\lambda(p) := \mathscr{U}(p/\lambda)/\mathscr{U}(p)$, $0 < p < 1$, $\lambda \geq 1$.

For the sake of completeness $\bar{q}_n$ has to be redefined on $(0, 1/n)$ by putting $\bar{q}_n(p) = 0$ there; similarly on $(1 - 1/n, 1)$. O'Reilly (1974) showed that as for the empirical process (cf. (2.2)), the integrability condition (2.3) with $h = h_\lambda$ is both necessary and sufficient for (3.6). We have thus arrived at the following theorem.

THEOREM 3.1 *Let F have a derivative f, which is positive and continuous on the support of F. Assume that the density-quantile function $\mathscr{U} = f \circ F^{-1}$ is nondecreasing (nonincreasing) in some neighborhood of zero (one), and that* (2.3) *is satisfied for each* $h \equiv h_\lambda$, $\lambda \geq 1$. *Then $q_n \to_{\mathscr{L}} B^\circ$ as $n \to \infty$.*

Analogous results may be obtained for the weighted quantile process. Our regularity conditions are somewhat weaker than those in Shorack (1972). Csörgő and Révész (1978) introduced the regularity condition

$$\sup_{x\in \operatorname{supp} F} F(x)(1 - F(x)) \left| \frac{f'(x)}{f^2(x)} \right| \leq \gamma < \infty. \tag{3.7}$$

However, this implies, e.g., that the function $p \to f(F^{-1}(p))/p^{2\gamma}$ is nonincreasing in a neighborhood of zero. Hence in the above considerations $f(F^{-1}(p))$ may be replaced by $p^{2\gamma}$. But $(p/\lambda)^{2\gamma}/p^{2\lambda} = \lambda^{-2\gamma} < \infty$ is bounded. Observe that, more generally, h_λ^{-1} is bounded from above (at zero) if $\mathscr{U}/g$ is nonincreasing in a neighborhood of zero for some regular varying function g.

Csörgő and Révész (1978) derived a somewhat stronger result than convergence in distribution. They considered the distance between the general and the uniform quantile process and obtained rates of uniform convergence under the stronger condition (3.7). However, their representation does not provide any information for determining confidence contours for statistics $T(q_n)$ in the finite sample case.

At the end of this chapter we shall derive some estimates for the relative error

$$\frac{F_n^{-1}(p) - F_n^{-1}(p')}{p - p'} \qquad \text{as } p - p' \to 0 \text{ and } n \to \infty.$$

As always suppose first that F_n comes from a uniform sample. In this case

$$\frac{F_n^{-1}(p) - F_n^{-1}(p')}{p - p'} = 1 + \frac{R_n(p) - R_n(p')}{p - p'} - \frac{\alpha_n(p) - \alpha_n(p')}{\sqrt{n}(p - p')},$$

where

$$R_n(p) := F_n^{-1}(p) + F_n(p) - 2p, \qquad 0 < p < 1.$$

Let $(a_n)_{\mathbb{N}}$ be any bandsequence. From (3.2) and (2.12) we get

$$\sup_{p-p'=a_n} \left| \frac{F_n^{-1}(p) - F_n^{-1}(p')}{p - p'} - 1 \right| = O\,(e_n) \qquad \mathbb{P}\text{-a.s.} \tag{3.8}$$

where

$$e_n = (\ln \ln n)^{1/4}(\ln n)^{1/2}n^{-3/4}a_n^{-1} + (\ln a_n^{-1}/na_n)^{-1/2}.$$

Suppose that $(a_n)_n$ is such that $e_n \to 0$ as $n \to \infty$. (3.8) then implies

$$\sup_{p-p'=a_n} \left| \frac{p - p'}{F_n^{-1}(p) - F_n^{-1}(p')} - 1 \right| = o(1) \qquad \mathbb{P} - \text{a.s.}$$

From this we get uniformly in $p - p' = a_n$, as $n \to \infty$:

$$\sqrt{na_n}\left(\frac{p-p'}{F_n^{-1}(p) - F_n^{-1}(p')} - 1\right) \sim a_n^{-1/2}[\alpha_n(p) - \alpha_n(p') + n^{1/2}(R_n(p') - R_n(p))]$$

(3.9)

$$= a_n^{-1/2}[\alpha_n(p) - \alpha_n(p') + O((\ln \ln n)^{1/4}(\ln n)^{1/2}n^{-1/4})].$$

Suppose that $n^{-\delta} = O(a_n)$ for some $\delta < \frac{1}{2}$. Then the last expression can be written

$$a_n^{-1/2}[\alpha_n(p) - \alpha_n(p')] + o(1). \tag{3.10}$$

Hence the left-hand side of (3.9) has asymptotically ($\mathbb{P}$-almost surely and in distribution) the same behavior as the modulus of the uniform empirical process.

The case of a general F with positive derivative f is again handled by considering the particular version (0.1), namely:

$$\frac{F_n^{-1}(p) - F_n^{-1}(p')}{p - p'} = \frac{1}{f(z)} \frac{\bar{F}_n^{-1}(p) - \bar{F}_n^{-1}(p')}{p - p'}, \tag{3.11}$$

where z is a suitable point between $\bar{F}_n^{-1}(p)$ and $\bar{F}_n^{-1}(p')$. Since $\bar{F}_n$ is the empirical distribution function of a uniform sample, the second factor may be estimated via (3.8). Furthermore, if f is bounded against zero on the support of F, we have that uniformly in $p - p' = a_n$, $F_n^{-1}(p) - F_n^{-1}(p')$ is of order $O(a_n)$ (provided the right-hand side of (3.8) converges to zero or is at least bounded). If f attains arbitrarily small values, then $F_n^{-1}(p) - F_n^{-1}(p')$ may be considerably larger than $p - p'$.

Remark 3.2. We did not aim at finding the most general growth conditions on $(a_n)_n$ for which $e_n \to 0$ as $n \to \infty$. This problem should be attacked along lines similar to Wellner (1978), where the convergence of $\bar{F}_n^{-1}(p)/p$ has been studied uniformly on intervals $(a_n, 1)$ under various growth conditions on a_n.

4. Estimation of A Density. For testing any hypothesis about the p-quantile $F^{-1}(p)$ of the underlying d.f. F, the asymptotic normality obtained in Theorem 3.1 suggests to consider the statistics

$$n^{1/2}(F_n^{-1}(p) - F^{-1}(p)).$$

However, to obtain proper confidence intervals it will be necessary to have some knowledge about the value of $f(F^{-1}(p))$. Therefore the problem arises of constructing an empirical estimate f_n of the true density f. For such an f_n the value of $f_n(F^{-1}(p))$ [or $f_n(F_n^{-1}(p))$] may then serve as an estimate of the variance of the normal limit distribution. Rosenblatt (1956) proposed to study a type of density estimate, which is obtained as a convolution between F_n and a properly scaled kernel function K:

$$f_n(t) := a_n^{-1} \int K\left(\frac{t-x}{a_n}\right) F_n(\mathrm{dx}), \qquad t \in \mathbb{R}.$$

Here, $(a_n)_n$ is a sequence of "window-widths" tending to zero as $n \to \infty$. In his fundamental paper Parzen (1962) showed that under some mild smoothness conditions on K (and f) $f_n(t)$ is a consistent estimator of $f(t)$. For example, if $K = 1_{[-1/2, 1/2)}$, we get the well-known "naive estimator"

$$f_n(t) = a_n^{-1}[F_n(t + \tfrac{1}{2} a_n) - F_n(t - \tfrac{1}{2} a_n)].$$

For sufficiently smooth f it is easy to see that

$$E_1^2 \equiv \mathrm{Var}[f_n(t)] = f(t)[na_n]^{-1}(1 + o(1))$$

and

$$E_2 \equiv \mathbb{E}(f_n(t)) - f(t) = O(a_n).$$

Hence f_n is consistent in quadratic mean if $na_n \to \infty$ (as well as $a_n \to 0$). Moreover, the random effect $f_n(t) - \mathbb{E}(f_n(t))$ becomes small if a_n is not too large, while the bias E_2 becomes small with a_n. The problem which arises is one of finding the optimal a_n which minimizes $E_1^2 + E_2^2$. This question has been extensively studied in the literature, even for more general kernel functions K. See for example the survey articles of Scott et al. (1977) and Wertz (1978). As to the almost sure convergence of f_n to f the existing results do not provide any satisfactory solution of this problem. Let us show below how Theorem 2.13 may be applied to get optimal window-widths, if f_n is the naive estimator. The case of a general kernel function K needs some further calculations and will be treated in a forthcoming paper (Stute (1982)).

We have to assume that the window-widths a_n form a bandsequence as defined in section 0. The deviation $f_n - f$ is measured on some central interval J of the real line. For each $\varepsilon > 0$, J_ε denotes the inner ϵ-parallel set of J. To avoid trivial statements, let $J_\varepsilon \neq \emptyset$ throughout. Theorem 2.13 then implies that, if

$$\bar{f}_n(t) := \mathbb{E}(f_n(t)) = a_n^{-1}[F(t + \tfrac{1}{2}a_n) - F(t - \tfrac{1}{2}a_n)],$$

one has

$$\text{(4.1)} \qquad \lim_{n\to\infty} \sqrt{\frac{na_n}{2 \ln a_n^{-1}}} \sup_{t\in J_\varepsilon} \frac{|f_n(t) - \bar{f}_n(t)|}{\sqrt{f(t)}} = 1 \qquad \mathbb{P} - \text{a.s.}$$

(put $\underline{c} = 1 = \bar{c}$ and $x_{u,t} = (u + t)/2$).

To obtain proper estimates for the bias E_2, one has to assume that higher derivatives of f exist. For example, if $f = F'$ is twice-continuously differentiable on J, then for all large enough $n \in \mathbb{N}$,

$$\sup_{t\in J_\varepsilon} |\bar{f}_n(t) - f(t)| = O(a_n^2),$$

provided that f'' is bounded. Hence E_2 is asymptotically negligible if $a_n^2 = o((\ln a_n^{-1}/na_n)^{1/2})$, i.e. when $na_n^5/\ln a_n^{-1} \to 0$.

In this case we have under the assumptions of Theorem 2.13

$$\text{(4.2)} \qquad \lim_{n\to\infty} \sqrt{\frac{na_n}{2 \ln a_n^{-1}}} \sup_{t\in J_\varepsilon} \frac{|f_n(t) - f(t)|}{\sqrt{f(t)}} = 1 \qquad \mathbb{P} - \text{a.s.}$$

If f is only once differentiable we get $E_2 = O(a_n)$. In this case the bias is asymptotically negligible whenever $na_n^3/\ln a_n^{-1} \to 0$.

Suppose now that f admits a third continuous derivative f'''. By Taylor's expansion we get

$$\text{(4.3)} \qquad \bar{f}_n(t) - f(t) = \frac{a_n^2}{2} f''(t) \int_{-1/2}^{1/2} y^2\, dy + O(a_n^3) \qquad \text{uniformly in } t \in J_\varepsilon$$

(provided f''' is bounded). Furthermore, if f is bounded and if $na_n^5/\ln a_n^{-1} \to \infty$, Theorem 2.15 implies

$$\lim_{n\to\infty} 2a_n^{-2} \sup_{t\in J_\varepsilon} |\bar{f}_n(t) - f(t)| = \sup_{t\in J_\varepsilon} |f''(t)| \int_{-1/2}^{1/2} y^2\, dy.$$

The optimal a_n is obtained if both $(f_n - \bar{f}_n)/\sqrt{f}$ and $(\bar{f}_n - f)/\sqrt{f}$ are of the same order. By (4.1) and (4.3) this is achieved by choosing a_n so as to minimize

$$a_n^2 \sup_{t\in J_\varepsilon} \frac{|f''(t)|}{\sqrt{f(t)}} \frac{1}{2} \int_{-1/2}^{1/2} y^2\, dy + \sqrt{\frac{2 \ln a_n^{-1}}{na_n}}.$$

One can easily see that for this a_n

$$a_n \sim \left[\frac{\ln n}{10n \sup_{t\in J_\varepsilon} |f''(t)|^2/f(t)\left(\int_{-1/2}^{1/2} y^2\,dy\right)^2}\right]^{1/5}, \qquad n \to \infty.$$

The dependence on f (and f'') makes it impossible to determine the exact order of a_n. Anyway, putting $a_n = (\ln n/n)^{1/5}$ will give us the appropriate rate at which the window-widths should converge to zero. For this choice of a_n we have $(f_n - f)/\sqrt{f} \to 0$ as $(\ln n/n)^{2/5}$ uniformly on J_ε.

It is known that the advantage of kernel type density estimation lies in a reduction of the bias of the resulting estimator. This property is not shared by the estimator which is undoubtedly the oldest, namely the histogram. One considers a sequence of points

$$\cdots < z_{-2}^{(n)} < z_{-1}^{(n)} < z_o^{(n)} < z_1^{(n)} < z_2^{(n)} < \cdots$$

on the real line, finite or not. Put $z_{-\infty}^{(n)} = -\infty$ and $z_\infty^{(n)} = \infty$. If J is the central portion of $\mathbb{R}$ where $f(t)$ should be estimated, and if we put $z_i \equiv z_i^{(n)}$, the corresponding histogram is defined by

$$f^*{}_n(t) := \begin{cases} \dfrac{F_n(z_{i+1}) - F_n(z_i)}{z_{i+1} - z_i} & \text{if } z_i \le t < z_{i+1} \text{ for some } i \in \mathbb{Z} \\ 0 & \text{otherwise}, \quad t \in J \end{cases}$$

In other words, $f_n^*(t)$ is a certain difference quotient of F_n which is determined from the location of t with respect to the grid $\{z_i^{(n)}\}$. To a certain extent this is similar to the definition of the naive kernel estimator. The only though important difference lies in the fact that in the latter case the points z_i are chosen symmetrically around t and therefore depend on t. Actually, this is responsible for the reduction of the bias.

In the classical setup the points $z_i = z_i^{(n)}$ are nonrandom and equidistant, say $z_{i+1} - z_i = a_n$, where $a_n \downarrow 0$. We have

$$f_n^*(t) - f(t) = n^{-1/2}\frac{\beta_n(z_{i+1}) - \beta_n(z_i)}{z_{i+1} - z_i} + \frac{1}{z_{i+1} - z_i}\int_{z_i}^{z_{i+1}} [f(x) - f(t)]\,dx,$$

$z_i \le t < z_{i+1}$. This shows that as in the case of kernel type estimation, the asymptotic properties of the histogram may be easily derived from the local behavior of the empirical process. If f is not constant on $[z_i, z_{i+1}]$ the best possible rate for the bias is of order $O(a_n)$. Hence up to a logarithmic term the rate of almost sure uniform convergence of $f_n^* \to f$ cannot be faster than $n^{-1/3}$. We only mention the following result which plays the same role as (4.2) for the kernel type estimators.

THEOREM 4.1 *Suppose that $f = F'$ is continuously differentiable on J with $0 < m \le f(x) \le M < \infty$ and $f'(x) \le M' < \infty$ for all $x \in J$. Let $(a_n)_n$ be any bandsequence such that $na_n^3/\ln a_n^{-1} \to 0$. Then if f_n^* is any histogram with nonrandom grid of length a_n, we have* $\mathbb{P}$ – a.s.

$$\lim_{n\to\infty} \sqrt{\frac{na_n}{2\ln a_n^{-1}}}\sup_{t\in J_\varepsilon}\frac{|f_n^*(t) - f(t)|}{\sqrt{f(t)}} = 1.$$

If f attains arbitrarily small values on J, it is still possible to obtain a (somewhat weaker) limit result, for which the right-hand side has the same form as in Theorem 2.15.

We are now going to study histograms corresponding to random grids. In most cases the points $z_i^{(n)}$ are taken from the set $\xi_1^{(n)} < \cdots < \xi_n^{(n)}$ of order statistics of the underlying sample $\xi_1, \cdots, \xi_n$. For example, one might consider a sequence $(k_n)_n$ of positive integers such that $a_n := k_n/n \to 0$ as $n \to \infty$. Put $z_i^{(n)} \equiv z_i := \xi_{ik_n}^{(n)}$, $i = 1, \cdots, \langle n/k_n \rangle$. By definition of f_n^*,

$$f_n^*(t) = \frac{F_n(z_{i+1}) - F_n(z_i)}{z_{i+1} - z_i}, \qquad z_i \le t < z_{i+1},$$

so that in this case

$$f_n^*(t) = \frac{a_n}{F_n^{-1}((i+1)a_n) - F_n^{-1}(ia_n)}.$$

For investigating the last term we may assume w.l.o.g. that F_n has the form considered in (0.1) and (3.11), respectively. We then have

$$f_n^*(t) = \mathcal{U}(z) \frac{a_n}{\bar{F}_n^{-1}((i+1)a_n) - \bar{F}_n^{-1}(ia_n)}, \tag{4.4}$$

where z is some point between $\bar{F}_n^{-1}(ia_n)$ and $\bar{F}_n^{-1}((i+1)a_n)$. Suppose that $(a_n)_n$ is some bandsequence with $n^{-\delta} = O(a_n)$, for some $\delta < ½$. Then from (3.9) and (3.10) we get uniformly in t ($\mathbb{P}$ – a.s.):

$$(na_n)^{1/2}\left(\frac{f_n^*(t)}{f(t)} - 1\right) \sim a_n^{-1/2}[\alpha_n((i+1)a_n) - \alpha_n(ia_n)] + o(1) + \frac{\mathcal{U}(z) - f(t)}{f(t)} O((na_n)^{1/2}). \tag{4.5}$$

The problem is one of finding sufficiently weak smoothness conditions on F (and f) such that the last term in (4.5) converges to zero uniformly in t. Parzen (1979) proposed to describe each possible continuous d.f. by its density-quantile function $\mathcal{U} = f \circ F^{-1}$. Recall (cf. Parzen (1979), page 162) that F has lower tail exponent $\alpha \in \mathbb{R}$ iff $\mathcal{U}(u) \sim u^\alpha$ as $u \to 0$, and upper tail exponent $\beta \in \mathbb{R}$ iff $\mathcal{U}(u) \sim (1-u)^\beta$ as $n \to 1$. Suppose that $\mathcal{U}$ has a continuous derivative $\mathcal{U}'$; then, e.g., $\alpha = \lim_{u \to 0} \dfrac{u\mathcal{U}'(u)}{\mathcal{U}(u)}$.

To estimate the last term in (4.5), note that by (3.8) $\mathbb{P}$-almost surely

$$|z - F(t)| \le |\bar{F}_n^{-1}((i+1)a_n) - \bar{F}_n^{-1}(ia_n)| = (1 + O(e_n))a_n$$

Hence

$$\frac{\mathcal{U}(z) - f(t)}{f(t)} = O(a_n)$$

uniformly in t on each finite subinterval of the real line. For small values of t we obtain for some u between z and $F(t)$

$$\frac{\mathcal{U}(z) - \mathcal{U}(F(t))}{\mathcal{U}(F(t))} = \frac{\mathcal{U}'(u)}{\mathcal{U}(F(t))}(z - F(t)) \sim \frac{\mathcal{U}'(u)}{\mathcal{U}(u)}\left(\frac{F(t)}{u}\right)^{-\alpha}(z - F(t)) \sim \frac{\alpha}{u}\left(\frac{F(t)}{u}\right)^{-\alpha}(z - F(t)).$$

Since

$$\frac{\bar{F}_n^{-1}(ia_n)}{\bar{F}_n^{-1}((i+1)a_n)} \le \frac{F(t)}{u} \le \frac{\bar{F}_n^{-1}((i+1)a_n)}{\bar{F}_n^{-1}(ia_n)}$$

it follows from Theorem 4 in Wellner (1978), that $\mathbb{P}$ – a.s.

$$F(t)/u = \mathcal{O}(1) \qquad \text{uniformly in } t$$

and therefore

$$\frac{\mathcal{U}(z) - f(t)}{f(t)} = O(a_n/u).$$

At this point we have to make one further restriction on the location of the point t. Suppose that $z_i \le t < z_{i+1}$ and i is such that $i_n \le i \le n - i_n$, $i_n/n \to 0$.
Then

$$1/u \le 1/\bar{F}_n^{-1}(i_n a_n) = O(1/i_n a_n) \qquad \mathbb{P} - \text{a.s.},$$

again by Theorem 4 in Wellner (1978). In summary, by (4.5),

$$(na_n)^{1/2}\left(\frac{f_n^*(t)}{f(t)} - 1\right) \sim a_n^{-1/2}[\alpha_n((i+1)a_n) - \alpha_n(ia_n)] + o(1) + O((na_n)^{1/2}/i_n)$$

$$= (na_n)^{1/2}[\bar{f}_n(F(t)) - 1] + o(1) + O((na_n)^{1/2}/i_n),$$

where $\bar{f}_n$ is the histogram of the uniform sample $\bar{\xi}_1, \cdots, \bar{\xi}_n$ with respect to the nonrandom grid $\bar{z}_i^{(n)} = ia_n$. We have thus obtained a representation of the general randomized histogram in terms of the uniform histogram with equidistant grid. For example, Theorem 2.10 immediately yields the following result.

THEOREM 4.2. *Suppose that F has upper and lower tail exponents. For each $n \in \mathbb{N}$ let $k_n \in \mathbb{N}$ be such that $a_n := k_n/n$ forms a bandsequence with $n^{-\delta} = \mathcal{O}(a_n)$ for some $\delta < \frac{1}{2}$. Let f_n^* denote the histogram with respect to the random grid $z_i^{(n)} = \xi_{ik_n}^{(n)}$. Then, if $i_n \in \mathbb{N}$ is such that $na_n/i_n^2 = \mathcal{O}(1)$,*

$$\lim_{n\to\infty} \sqrt{\frac{na_n}{2 \ln a_n^{-1}}}\, \sup_{\xi_{i_nk_n}^{(n)} \le t \le \xi_{n-i_nk_n}^{(n)}} \left|\frac{f_n^*(t) - f(t)}{f(t)}\right| = 1 \qquad \mathbb{P} - \text{a.s.}$$

The essence of the result is that better rates of convergence may be obtained at the price, that the deviation between f_n^* and f is measured only between $\xi_{i_nk_n}^{(n)}$ and $\xi_{n-i_nk_n}^{(n)}$ (of course, $i_nk_n/n \to 0$). This important property is not shared by the ordinary histogram considered in Theorem 4.1. In the same way it is also possible to derive distributional results for f_n^* from corresponding results for $\bar{f}_n$ (cf., e.g., Bickel and Rosenblatt (1973)).

It is an important fact that similar estimates are also valid if, for each $t \in \mathbb{R}$, the points $z_i \le t < z_{i+1}$ are taken from the set of order statistics $\xi_1^{(n)} < \cdots < \xi_n^{(n)}$ in such a way, that z_i and z_{i+1} have a spacing k_n also in this case (i.e. $z_i = \xi_j^{(n)}$ and $z_{i+1} = \xi_{j+k_n}^{(n)}$ for some j), but may differ from t to t, depending on the location of t. Such histogram type estimators have been investigated, among others, by van Ryzin (1973). An attractive feature of (4.5) is that one is led to choose z_i and z_{i+1} so that $|F^{-1}(z) - t|$ and hence $|\mathcal{U}(z) - f(t)|$ become small. This is achieved by choosing j so as to minimize the set of values $\xi_{i+k_n}^{(n)} - \xi_i^{(n)}$ with $\xi_i^{(n)} \le t < \xi_{i+k_n}^{(n)}$. We have thus arrived at the so-called "nearest neighbor density estimator" which was first proposed by Loftsgaarden and Quesenberry (1965). Moore and Henrichon (1969) investigated the almost sure uniform consistency of such estimators. We also refer to Moore and Yackel (1977), and Devroye and Wagner (1977). Theorem 4.2 also applies in this case and is, as far as we know, the first estimate of this form.

REFERENCES

[1] BAHADUR, R. R. (1969). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.

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

[3] BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

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

[5] BOHMAN, H. (1963). Two inequalities for Poisson distributions. *Skand. Aktuarietidskr.* **46**, 47–52.

[6] BREIMAN, L. (1968). *Probability*. Addison-Wesley. Reading.
[7] BRILLINGER, D. R. (1969). The asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* **75,** 545–547.
[8] CIBISOV, D. M. (1964). Some theorems on the limiting behavior of the empirical distribution function. *Selected Translations in Math. Statist. and Prob.* **6,** 147–156 (1966).
[9] CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximation of the quantile process. *Ann. Statist.* **6,** 882–894.
[10] DEVROYE, L. P. and WAGNER, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5,** 536–540.
[11] GAENSSLER, P. and STUTE, W. (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probability* **7,** 193–243.
[12] JAESCHKE, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7,** 108–115.
[13] KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38,** 1323–1342.
[14] KIEFER, J. (1970). Deviations between the sample quantile process and the sample d.f. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 299–319. Cambridge Univ. Press. London.
[15] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df. *I. Z. Wahrsch. verw. Gebiete* **32,** 111–132.
[16] LOFTSGAARDEN, D. O. and QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36,** 1049–1051.
[17] MOORE, D. S. and HENRICHON, E. G. (1969). Uniform consistency of some estimates of a density function. *Ann. Math. Statist.* **40,** 1499–1502.
[18] MOORE, D. S. and YACKEL, J. W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.* **5,** 143–154.
[19] O'REILLY, N. E. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability* **2,** 642–651.
[20] PARTHASARATHY, K. R. (1967). Probability Measures on Metric Spaces. *Academic Press.* New York.
[21] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33,** 1065–1076.
[22] PARZEN, E. (1979). Density quantile estimation approach to statistical data modelling. In *Smoothing Techniques for Curve Estimation.* Lect. Notes in Math. 757 (Th. Gasser and M. Rosenblatt ed.'s) 155–180. Springer. Berlin.
[23] PYKE, R. and ROOT, D. H. (1968). An application of stopping times to obtain weak convergence. Technical Report No. 16. Univ. of Washington.
[24] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39,** 755–771.
[25] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27,** 832–837.
[26] SCOTT, D. W., TAPIA, R. A. and THOMPSON, J. R. (1977). Kernel density estimation revisited. *Nonlinear Anal.,* Theor., Math. Appl. **1,** 339–372.
[27] SHORACK, G. R. (1972). Convergence of quantile and spacing processes with applications. *Ann. Math. Statist.* **43,** 1400–1411.
[28] SHORACK, G. R. and WELLNER, J. A. (1982). Limit theorems and inequalities for the uniform empirical processes indexed by intervals. *Ann. Probability.* To appear.
[29] STUTE, W. (1982). A law of the logarithm for kernel density estimators. To appear.
[30] TAKÁCS, L. (1967). Combinatorial Methods in the Theory of Stochastic Processes. Wiley. New York.
[31] VAN RYZIN, J. (1973). A histogram method of density estimation. *Comm. Statist.* **2,** 493–506.
[32] WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. verw. Gebiete* **45,** 73–88.
[33] WERTZ, W. (1978). Statistical density estimation. A survey. Vandenhoeck und Ruprecht. Göttingen.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MUNICH
THERESIENSTRAßE 39
D-8000 MUNICH 2