The Annals of Statistics
1992, Vol. 20, No. 2, 1062–1078

GENERALIZED QUANTILE PROCESSES

BY JOHN H. J. EINMAHL[1] AND DAVID M. MASON[2]

Eindhoven University of Technology and University of Delaware

For random vectors taking values in $\mathbb{R}^d$ we introduce a notion of multivariate quantiles defined in terms of a class of sets and study an associated process which we call the generalized quantile process. This process specializes to the well known univariate quantile process. We obtain functional central limit theorems for our generalized quantile process and show that both Gaussian and non-Gaussian limiting processes can arise. A number of interesting example are included.

1. Introduction and main results. The classical real-valued quantile process continues to have wide ranging applications in statistics and probability. Refer, in particular, to the books of Csörgő and Révész (1981), Csörgő (1983) and Shorack and Wellner (1986). The purpose of this paper is to introduce a notion of multivariate quantiles defined by means of a class of sets and to investigate functional central limit theorems for the associated quantile process. This process will be seen to be a natural generalization of the classical real-valued quantile process. It should prove to be a useful new tool to deal with inference about multivariate data, for instance in the construction of goodness of fit tests, especially generalized *Q-Q* plots.

To begin with, let $X_1, \dots, X_n$, $n \geq 1$, be independent random vectors taking values in $\mathbb{R}^d$, $d \geq 1$, with common distribution function F. Further, let $\mathbb{A}$ be a subset of the Borel sets $\mathbb{B}$ on $\mathbb{R}^d$ and introduce the pseudometric d_0 defined on $\mathbb{B}$ by

$$d_0(B_1, B_2) = P(B_1 \vartriangle B_2), \quad \text{for } B_1, B_2 \in \mathbb{B},$$

where P is the probability measure on $(\mathbb{R}^d, \mathbb{B})$ pertaining to F. Define the empirical measure P_n on $\mathbb{B}$ by

$$P_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i), \qquad B \in \mathbb{B},$$

where 1_B denotes the indicator function. Now let λ be a real-valued function defined on $\mathbb{A}$. By the quantile function based on P, λ and $\mathbb{A}$ we shall mean the function

$$U(t) = \inf\{\lambda(A) : P(A) \geq t,\, A \in \mathbb{A}\}, \qquad 0 < t < 1, \tag{1.1}$$

Received May 1990; revised July 1991.

[1]Research performed at the University of Limburg, Maastricht.

[2]Research partially supported by the Nederlandse organisatie voor wetenschappelijk onderzoek while the author was visiting the University of Limburg and NSF Grant DMS-88-0329.

AMS 1980 *subject classifications*. 60F05, 62E20, 62G30.

Key words and phrases. Empirical measure, generalized quantile process, central limit theorem.

and by the empirical quantile function,

$$(1.2)\quad U_n(t) = \inf\{\lambda(A): P_n(A) \geq t,\ A \in \mathbb{A}\}, \qquad 0 < t < 1\ (\inf \phi = \infty).$$

Often, a natural choice of λ is to be Lebesgue measure λ_d on $\mathbb{R}^d$. In this case $U_n(t)$ is roughly the "volume" of the smallest set in $\mathbb{A}$ that contains at least fraction t of the data points. A definition similar to (1.2) was used to define minimal multivariate spacings in Deheuvels, Einmahl, Mason and Ruymgaart (1988); this idea is also latent in Pyke (1984) and Rousseeuw (1985).

Depending on the smoothness of the sets in $\mathbb{A}$, other choices for λ that may be feasible are the length of the perimeter of a set, its diameter or a (probability) measure evaluated at the set; in short, whatever may be reasonable in a given setup. For instance, for data on a circle a natural choice for $\mathbb{A}$ is the set of all arcs on the circle and λ is the arc length.

By selecting λ and $\mathbb{A}$ appropriately, our generalized quantile function can describe various features of the underlying distribution function F. When $d = 1$ by choosing $\mathbb{A} = \{(-\infty, x]: x \in \mathbb{R})\}$ and $\lambda((-\infty, x]) = x$, we get from (1.1) and (1.2) the definitions of the quantile and empirical quantile functions in the classical real-valued case; by setting $\mathbb{A} = \{[a, b]:\ -\infty < a < b < \infty\}$ and $\lambda([a, b]) = b - a$, $U_n(t)$ becomes the length of the shortest interval containing fraction t of the data (shortt for short) as studied by Grübel (1988) in connection with robust scale estimation. For further details along this line, see Rousseeuw and Leroy (1988). The shortt along with some true multivariate examples is considered in more detail in Section 2.

In order to establish the above mentioned functional central limit theorems, we must impose some regularity conditions on the class $\mathbb{A}$ and the function λ. First we require that:

(C_1) λ is continuous on $\mathbb{A}$ with respect to the pseudometric d_0 and λ and $\mathbb{A}$ are such that the U_n, $n \geq 1$, are finite valued on $(0, 1)$ almost surely.

(C_2) There exists a countable subclass $\mathbb{D}$ of $\mathbb{A}$ such that for any $A \in \mathbb{A}$ there is a sequence $\{D_n, n \geq 1\}$ in $\mathbb{D}$ with $1_{D_n}(x) \to 1_A(x)$ for all $x \in \mathbb{R}^d$.

[Assumption (C_2) is assumption (SE) in Lemma 20 of Gaenssler (1983), page 108; we impse it here to avoid measurability problems.]

Let $\mathbb{A}^* = \mathbb{A} \cup \{\mathbb{R}^d, \phi\}$. Following Dudley (1978) and Gaenssler (1983), Chapter 4, let

$$S_0 = \{\psi: \mathbb{A}^* \to \mathbb{R}: \psi \text{ is bounded and uniformly } d_0\text{-continuous}\}$$

and set, with δ_x denoting the unit mass at x,

$$S = \left\{\psi = \psi_1 + \psi_2: \psi_1 \in S_0 \text{ and } \psi_2 = \sum_{i=1}^{k} a_i \delta_{x_i}\right.$$
$$\left.\text{for some } a_i \in \mathbb{R},\ x_i \in \mathbb{R}^d,\ k \in \mathbb{N}\right\}.$$

Finally, equip S with the supremum metric ρ:

$$\rho(\psi', \psi'') := \sup_{A \in \mathbb{A}^*} |\psi'(A) - \psi''(A)|, \quad \text{for } \psi', \psi'' \in S.$$

Define for each $n \geq 1$ the empirical process indexed by $\mathbb{A}^*$ to be

$$\alpha_n(A) = n^{1/2}\{P_n(A) - P(A)\}, \qquad A \in \mathbb{A}^*.$$

By Lemma 20 of Gaenssler (1983), assumption (C_2) implies that α_n is a random element in S, where S is provided with the σ algebra generated by the open balls. We shall assume that:

(C_3) α_n converges weakly in (S, ρ) [in the sense of Dudley (1978)] to B_P, a bounded, mean zero Gaussian process indexed by $\mathbb{A}^*$, uniformly continuous in d_0 on $\mathbb{A}^*$, with covariance function $P(A_1 \cap A_2) - P(A_1)P(A_2)$, $A_1, A_2 \in \mathbb{A}^*$.

Assumption (C_3) implies that $\mathbb{A}$ is totally bounded and, furthermore, by the Skorohod–Dudley–Wichura representation theorem [e.g., Gaenssler (1983), page 82] there exists a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ carrying a version $\tilde{B}_P$ of B_P and a sequence of versions $\tilde{\alpha}_n$ of α_n such that

$$\sup_{A \in \mathbb{A}^*} \left|\tilde{\alpha}_n(A) - \tilde{B}_P(A)\right| \to 0 \quad \text{a.s.} \tag{1.3}$$

Henceforth, without confusion, we will drop the tildas from the notation.

We shall need some additional technical assumptions on $\mathbb{A}$, P and λ, which in the classical real-valued case are trivially fulfilled under standard smoothness assumptions [see Shorack and Wellner (1986)]. These are the following:

(C_4) For all $A \in \mathbb{A}$, $0 < P(A) < 1$.

Set $T_0 = \{\phi\}$, $T_1 = \{\mathbb{R}^d\}$ and

$$T_t = \{A \in \mathbb{A}: \lambda(A) = U(t), P(A) = t\} \quad \text{for } 0 < t < 1. \tag{1.4}$$

We assume

(C_5) $T_t \neq \phi$ for all $0 < t < 1$

and

(C_6) for every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $0 \leq t_1, t_2 \leq 1$ with $|t_1 - t_2| < \delta$ and $A_1 \in T_{t_1}$ there is an $A_2 \in T_{t_2}$ with $d_0(A_1, A_2) < \varepsilon$.

For any $0 \leq t \leq 1$ let

$$B(t) = \sup_{A \in T_t} B_P(A). \tag{1.5}$$

We note that if T_t contains at least two sets A_1 and A_2 with $P(A_1 \vartriangle A_2) > 0$, then $B(t)$ is not a normal random variable and $EB(t) > 0$. Observe that (C_4) and (C_6) imply that B is continuous on $[0, 1]$ with $B(0) = B(1) = 0$ almost surely. Also from assumptions (C_1), (C_2) and (C_5) it is straightforward to infer that $U_n(t)$ is measurable for each $t \in (0, 1)$, and, moreover, since $U_n(t)$ is constant on $((i-1)/n, i/n]$ for $i = 1, \dots, n-1$ and on $(1 - 1/n, 1)$, we see that for each $0 < a < b < 1$, $U_n(\cdot)$ is a random element in $D_L[a, b]$, the space of left continuous functions with right limits on $[a, b]$. Furthermore, (C_1) and (C_6) imply that U is continuous on $(0, 1)$.

Finally we require two more assumptions. Let

$$L = \lim_{t \downarrow 0} U(t) \quad \text{and} \quad R = \lim_{t \uparrow 1} U(t). \tag{1.6}$$

Assume:

(C_7) U is strictly increasing on $(0, 1)$ with inverse $H = U^{-1}$ having a continuous derivative h on (L, R). [Note $H(x) = \sup\{P(A)\colon \lambda(A) \le x\}$, $L < x < R$.]

(C_8) For every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $A \in \mathbb{A}$ satisfies $0 < t - \delta < P(A) < t < 1$ and $\lambda(A) < U(t)$ there is an $A' \in T_{H(\lambda(A))}$ such that $d_0(A, A') < \varepsilon$.

Denote $g = h \circ U$ and consider the *generalized quantile process*

$$\beta_n(t) = g(t) n^{1/2}(U_n(t) - U(t)), \qquad 0 < t < 1. \tag{1.7}$$

We are now prepared to state our first theorem.

THEOREM 1.1. *Under assumptions* (C_1)–(C_8) *for all* $0 < a < b < 1$, *on the probability space of* (1.3),

$$\sup_{a \le t \le b} |\beta_n(t) + B(t)| \to 0 \quad a.s.\ as\ n \to \infty. \tag{1.8}$$

In the classical real-valued case, this theorem reduces to the well known Hájek–Bickel result [see Theorem 1 in Shorack and Wellner (1986), pages 640–641]. Our next theorem shows that under additional assumptions the convergence in (1.8) can be extended to the entire interval $(0, 1)$.

THEOREM 1.2. *In addition to the assumptions of Theorem* 1.1, *suppose* λ *is nonnegative*, $0 < h(L+) < \infty$, h *is nonincreasing in a left neighborhood of* R. *If* $\lim_{t \uparrow R} h(t) = 0$, *assume moreover that* h' *exists in a left neighborhood of* R *and for some* $0 < M < \infty$,

$$\limsup_{t \uparrow R} (1 - H(t)) |h'(t)| (h(t))^{-2} < M. \tag{1.9}$$

Then on the probability space of (1.3),

$$\sup_{0 < t < 1} |\beta_n(t) + B(t)| \to_{\mathbb{P}} 0 \quad as\ n \to \infty. \tag{1.10}$$

We note here that condition (1.9) is a right-tailed version of the Csörgő and Révész (1978) condition. Refer to Shorack and Wellner (1986) for a detailed account of conditions under which (1.10) holds for the classical real-valued quantile process.

REMARK 1.1. It is clear from our setup that the sample space $\mathbb{R}^d$ can be replaced by a general metric space. Moreover, it is easily seen that the only feature of weak convergence that we use in our proofs is statement (1.3). For the latest word on representations as in (1.3), refer to Dudley (1985).

Discussion. Our notion of multivariate quantiles is not intended to provide a recipe for ordering data in $\mathbb{R}^d$, but rather to offer a flexible technique to summarize properties of multidimensional data by means of a univariate quantile type of function. It is, however, roughly related to a number of methods that have been suggested for ordering (multidimensional) data, two of the most prominent of these being the idea of peeling the convex hull of the set of data points $X_1, \dots, X_n$ and that of ordering by means of an auxiliary function ψ, that is, $x \leq_\psi y$ if and only if $\psi(x) \leq \psi(y)$. The first method was originally suggested by Tukey (1975) and the second came out of the discussion by Plackett (1976) of Barnett's (1976) stimulating paper on multivariate ordering; see also Chapter 2 of Reiss (1989). [For other approaches to multivariate quantiles refer to Pyke (1975, 1985) and Eddy (1985).] Our quantiles are related to the first method in that a class of sets is used in the definition and to the second method, as pointed out by one of the referees, in the following way: Choose $\mathbb{A}$ to be the class of all sets of the form

$$A_r = \{x \colon \psi(x) \leq r\}, \qquad -\infty < r < \infty,$$

and set $\lambda(A_r) = r$. In this case $U_n((i/n)-)$, $i = 1, \dots, n$, become the order statistics of $\psi(X_i)$, $i = 1, \dots, n$.

This referee also remarked that one can further generalize quantile functions of this form by regarding them as functions indexed by both t and ψ. For example, let $\psi_j(x) = x_j$; that is, the projection on the jth coordinate, $1 \leq j \leq d$, of x. Also, for each $1 \leq j \leq d$, let $\mathbb{A}_j$ be the class of sets

$$A_{r,j} = \{x \colon x_j \leq r\}, \qquad r \in \mathbb{R},$$

and set $\lambda_j(A_{r,j}) = r$. Then for each coordinate j we get an empirical quantile function $U_{n,j}$ and $U_{n,j}((i/n)-)$, $i = 1, \dots, n$, become the ordered values of the jth coordinate of the X_i, $i = 1, \dots, n$. For instance, $(U_{n,1}(1-), \dots, U_{n,d}(1-))$ is the d-variate maximum studied in multivariate extreme value theory [see Resnick (1987)].

2. Examples.

The following two propositions will be very useful in presenting some of our examples.

PROPOSITION 2.1. *Assume that F is a distribution function on $\mathbb{R}^d$, $d \geq 1$, such that*

(2.1) *F has a continuous density f on $\mathbb{R}^d$,*

(2.2) *for all $0 \leq c < \infty$, $\lambda_d\{x\colon f(x) = c\} = 0$, with λ_d being Lebesgue measure on $\mathbb{R}^d$.*

Then for any $0 < t < 1$ *there exists a* $c(t) > 0$ *such that with* $C_t := \{x\colon f(x) \geq c(t)\}$,

$$P(C_t) = t, \tag{2.3}$$

(2.4) C_t *minimizes* $\lambda_d(C)$ *among all Lebesgue measurable sets* C *such that* $P(C) = t$ *and, moreover, if* C *is Lebesgue measurable,* $\lambda_d(C) = \lambda_d(C_t)$ *and* $P(C) = t$, *then* $\lambda_d(C \triangle C_t) = 0$.

Proof. Assertion (2.3) is an easy consequence of (2.1) and (2.2). The second part of the proposition follows directly from the generalized Neyman–Pearson lemma given in Lehmann [(1986), page 96] with $\mu = \lambda_d$, $m = 1$, $f_1 = f$ and $f_2 = -1$. □

Remark 2.1. Proposition 2.1 says that if λ is chosen to be λ_d, F satisfies (2.1) and (2.2) and for each $0 < t < 1$ there is an $A_t \in \mathbb{A}$ such that $\lambda(A_t \triangle C_t) = 0$, then A_t is the essentially unique $A \in T_t$ determining $U(t)$.

In order to state our next proposition, we need to introduce the following notation and assumptions: Let F be a distribution function on $\mathbb{R}^d$, $d \geq 1$, with a density f (w.r.t. Lebesgue measure λ_d) such that:

A1. conditions (2.1) and (2.2) hold;
A2. $f(Tx) = f(x)$ for every $x \in \mathbb{R}^d$ and orthogonal transform T on $\mathbb{R}^d$;
A3. for any $c > 0$, $\{x\colon f(x) \geq c\}$ is either empty or a closed centrally symmetric, convex set.

Assumptions A2 and A3 imply that whenever $\{x\colon f(x) \geq c\}$ is nonempty for some $c > 0$, it is necessarily a closed ball with center 0, from which it easily follows that

$$\sup_{x \in \mathbb{R}^d} f(x) =: D < \infty \tag{2.5}$$

and

$$\{\{x\colon f(x) \geq c\} : 0 < c < D\} = \{rE : 0 < r < \infty\}, \tag{2.6}$$

where E is the closed ball with center 0 and radius 1.

Distributions which satisfy A2 are called spherical. For a discussion of properties of spherical distributions along with many examples, see the book of Muirhead [(1982), pages 32–40].

Let $\mathbb{T}$ denote the set of all orthogonal transforms $T\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\mathbb{S}$ be the set of all scale transforms on $\mathbb{R}^d$; that is, each $S \in \mathbb{S}$ is of the form

$$S = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_d \end{pmatrix}, \quad \text{where } a_i > 0,\ i = 1, \dots, d.$$

For each $S \in \mathbb{S}$, write $|S| = \prod_{i=1}^d a_i$ = determinant of S. Consider the class of all closed ellipsoids

$$\mathbb{E} = \{TSE + z: T \in \mathbb{T},\ S \in \mathbb{S},\ z \in \mathbb{R}^d\}.$$

Under the preceding assumptions we can show, using Theorem 1 of Anderson (1955), that for each $T \in \mathbb{T}$, $S \in \mathbb{S}$ and $z \in \mathbb{R}^d$, $z \neq 0$,

$$P(TSE + kz) \text{ is strictly decreasing in } k \in [0, \infty). \tag{2.7}$$

For any $S \in \mathbb{S}$ and $z \in \mathbb{R}^d$ write $E_S(z) = SE + z$ and for any $0 < r < \infty$ denote by $B_r = rE$ the closed ball of radius r centered at 0.

PROPOSITION 2.2. *Under assumptions* A1–A3, *for all* $\varepsilon > 0$, *there exists a* $\delta > 0$ *such that whenever* $A \in \mathbb{E}$ *satisfies*

$$P(B_r) - P(A) < \delta, \tag{2.8}$$

where $A = TSE + z$ *for some* $T \in \mathbb{T}$, $S \in \mathbb{S}$ *and* $z \in \mathbb{R}^d$, *and* $r = |S|^{1/d}$, *then*

$$P(B_r \vartriangle A) < \varepsilon. \tag{2.9}$$

PROOF. First by A2 we can restrict our consideration to the subclass $\{E_S(z): S \in \mathbb{S},\ z \in \mathbb{R}^d\}$ of $\mathbb{E}$.

For $R \geq 1$ and $\eta > 0$ define with r as in (2.8),

$$\begin{aligned}\psi_1(R, \eta) = \sup\{P(B_r \vartriangle E_S(z)): |z| \leq \eta,\ R^{-1} \leq |S|^{1/d} \leq R,\\ |S|^{1/d} - \min(a_1, \dots, a_d) \leq \eta\},\end{aligned} \tag{2.10}$$

$$\begin{aligned}\psi_2(R, \eta) = \inf\{P(B_r) - P(E_S(z)): |z| > \eta,\ R^{-1} \leq |S|^{1/d} \leq R,\\ |S|^{1/d} - \min(a_1, \dots, a_d) \leq \eta\},\end{aligned} \tag{2.11}$$

$$\begin{aligned}\psi_3(R, \eta) = \inf\{P(B_r) - P(E_S(0)): R^{-1} \leq |S|^{1/d} \leq R,\\ |S|^{1/d} - \min(a_1, \dots, a_d) > \eta\}.\end{aligned} \tag{2.12}$$

It is readily verified using Proposition 2.1, A1–A3 and (2.7) combined with continuity that for each $R \geq 1$ and $i = 1, 2, 3$,

$$\psi_i(R, \eta) > 0 \quad \text{and} \quad \psi_i(R, \eta) \to 0 \quad \text{as } \eta \downarrow 0. \tag{2.13}$$

Set

$$\delta(R, \eta) = \min(\psi_2(R, \eta), \psi_3(R, \eta)) \tag{2.14}$$

and

$$\varepsilon(R, \eta) = 2P(B_{R^{-1}}) + 2(1 - P(B_R)) + \psi_1(R, \eta) + \delta(R, \eta). \tag{2.15}$$

We claim that for all $R \geq 1$ and $\eta > 0$ whenever, with $r = |S|^{1/d}$,

$$P(B_r) - P(E_S(z)) < \delta(R, \eta), \tag{2.16}$$

then

$$P(B_r \vartriangle E_S(z)) < \varepsilon(R, \eta). \tag{2.17}$$

To see this we must consider a number of cases.

CASE 1. $R^{-1} \leq r \leq R$, $|z| \leq \eta$, $|S|^{1/d} - \min(a_1, \dots, a_d) \leq \eta$. In this case (2.17) holds by definition of $\psi_1(R, \eta)$.

CASE 2. $R^{-1} \leq r \leq R$, $|z| > \eta$, $|S|^{1/d} - \min(a_1, \dots, a_d) \leq \eta$. By definition of $\psi_2(R, \eta)$, (2.16) cannot hold in this case.

CASE 3. $R^{-1} \leq r \leq R$, $|S|^{1/d} - \min(a_1, \dots, a_d) > \eta$. In this case by (2.7),

$$P(B_r) - P(E_S(z)) \geq P(B_r) - P(E_S(0)) \geq \psi_3(R, \eta),$$

so (2.16) is not satisfied.

CASE 4. $R^{-1} > |S|^{1/d}$. Here we have trivially

$$P(B_r \vartriangle E_S(z)) \leq 2P(B_{R^{-1}}),$$

which implies (2.17) by definition of $\varepsilon(R, \eta)$.

CASE 5. $|S|^{1/d} > R$. By elementary bounds,

$$P(B_r \vartriangle E_S(z)) \leq 2(1 - P(B_R)) + P(B_r) - P(E_S(z)).$$

Thus (2.16) implies (2.17).

Since by (2.13), for each $R \geq 1$, both $\delta(R, \eta) \to 0$ and

$$\varepsilon(R, \eta) \to 2P(B_{R^{-1}}) + 2(1 - P(B_R)), \quad \text{as } \eta \downarrow 0, \tag{2.18}$$

and the right side of (2.18) converges to zero as $R \uparrow \infty$, a routine argument now yields the assertion of Proposition 2.2. □

EXAMPLE 1. Let F be a distribution function on $\mathbb{R}^d$ satisfying Assumptions A1–A3 of Proposition 2.2 and set $\mathbb{A} = \mathbb{E}$ and $\lambda = \lambda_d$. For such an F, $\mathbb{A}$ and λ it is immediately obvious that assumptions (C_1), (C_2), (C_4) and (C_6) of Theorem 1.1 hold. To verify (C_3) we first recall the well known fact that the class of closed ellipsoids forms a Vapnik–Cervonenkis class and then combine this with (C_2), which implies by the Pollard (1982) central limit theorem that (C_3) is fulfilled. Assumption (C_7) follows by elementary analysis. In fact, here the function h is strictly decreasing, $L = 0$ and $h(0+) = D$, with D as in (2.5). Finally, after a little reflection, it follows that in this case (C_8) is a consequence of Proposition 2.2 and (C_7).

Choose $T \in \mathbb{T}$, $S \in \mathbb{S}$ and $z \in \mathbb{R}^d$. Then a little computation shows for $X_1, \dots, X_n$ i.i.d. F and $Y_i = TSX_i + z$, $i = 1, \dots, n$, that for all $0 < t < 1$, $\beta_n^Y(t) = \beta_n^X(t)$, where β_n^X and β_n^Y are the generalized quantile processes corresponding to $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, respectively. Thus, from the above, we see by Theorem 1.1 that (1.8) holds for β_n^Y with $0 < a < b < 1$ being arbitrary. Moreover, whenever (1.9) is satisfied, (1.10) holds. It is readily checked that the process B which arises is the *standard Brownian bridge*. [Use Proposition 2.1 and property (2.6).] For related work, see Davies (1987).

An aside (A natural way to choose the class $\mathbb{A}$). Assume that $Y_1, \dots, Y_n$ are i.i.d. $f(x;\theta)$, where $\{f(x;\theta): \theta \in \Theta\}$ is an exponential family of densities on $\mathbb{R}^d$ of the form

$$C(\theta)\exp\left(\sum_{j=1}^{k} \theta_j T_j(x)\right) h(x),$$

where Θ contains an open subset of $\mathbb{R}^k$ and $T_1, \dots, T_k$ are linearly independent real-valued measurable functions on $\mathbb{R}^d$ and h is a nonnegative measurable function on $\mathbb{R}^d$.

Since the class of functions

$$\sum_{j=1}^{k} \theta_j T_j, \qquad \theta = (\theta_1, \dots, \theta_k) \in \Theta,$$

is finite dimensional, one can show using Theorem 7.2 in Dudley (1978) that the class of sets

$$\{\{x: f(x;\theta) > c\}, \theta \in \Theta, c \geq 0\} =: \mathbb{A}^*$$

is a Vapnik–Cervonenkis (VC) class. If a little continuity is assumed, then the necessary weak convergence assumptions hold for the class $\mathbb{A}^*$, since VC plus measurability implies "weak convergence."

In the case when the $f(x, \theta)$ are the densities of the d-dimensional normal random vectors, the ellipsoids are generated in this way.

EXAMPLE 1(a) (Multivariate normal). It is simple to verify that all nondegenerate multivariate normal distribution functions are included within the setup of Example 1 and that they satisfy (1.9). Specializing to the case when $Y_1, \dots, Y_n$ are i.i.d. bivariate normal random vectors with mean (μ_1, μ_2) and variance–covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $-1 < \rho < 1$ and $\sigma_1, \sigma_2 > 0$, we obtain on the probability space of (1.3) as $n \to \infty$,

$$\sup_{0<t<1} \left| \frac{(1-t)}{\tau} n^{1/2} \left\{ U_n(t) - \tau \log\left(\frac{1}{1-t}\right) \right\} + B(t) \right| \to_{\mathbb{P}} 0,$$

with $\tau = 2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}$ and B a standard Brownian bridge.

EXAMPLE 1(b) (Length of shortt). Let F be a distribution function on $\mathbb{R}$ with density f which is positive, continuous, strictly increasing on $(-\infty, 0]$ and symmetric about zero. Such an F satisfies conditions A1–A3 of Proposition 2.2. Here the class $\mathbb{E} = \{[a, b]: -\infty < a < b < \infty\}$ and choosing $\lambda = \lambda_1$, we get $U(t) = 2F^{-1}((1+t)/2)$ and $g(t) = f(F^{-1}((1+t)/2))$. For a sample $Y_1, \dots, Y_n$ i.i.d. $F((\cdot - \mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$. $U_n(t)$ is easily seen to be the length of the smallest closed interval that contains at least fraction t of the observations. This we call the length of the shortt. We have in this case that for all $0 < a < b < 1$, $\mu \in \mathbb{R}$ and $\sigma > 0$ on the probability space of (1.3) as $n \to \infty$,

$$\sup_{a \le t \le b} \left| \sigma^{-1} g(t) n^{1/2} (U_n(t) - \sigma U(t)) + B(t) \right| \to 0 \quad \text{a.s.}$$

This result agrees with Theorem 3 of Grübel (1988). Note that he requires slightly more regularity conditions on the density. We should mention here that Kim and Pollard (1990) have previously remarked that Grübel's result can be obtained by means of abstract empirical process methods. Our next example extends Example 1(b).

EXAMPLE 2. Let F be a distribution function with a density which is positive and continuous on (α, β) where $-\infty \le \alpha < \beta \le \infty$ and zero elsewhere. Furthermore, assume that for some $t_0 \in (\alpha, \beta)$, f is nondecreasing on $(\alpha, t_0]$ and nonincreasing on $[t_0, \beta)$. Let $\mathbb{A} = \{[a, b]: \alpha < a < b < \beta\}$ and $\lambda = \lambda_1$. Now (C_1)–(C_4) are trivially fulfilled. For the verification of (C_5)–(C_8) the following observation is crucial: For every $A \in T_t$, $0 < t < 1$, there exists a $c(t)$ such that

$$\{x: f(x) > c(t)\} \subset A \subset \{x: f(x) \ge c(t)\}$$

and, moreover, $c(t) = h(t)$, where h is as in (C_7). From this, (C_5)–(C_7) are readily derived.

Condition (C_8) can be established using an adaption of the proof of Proposition 2.2. We sketch here the modifications. For every $0 < t < 1$, let $A_1(t)$ be the unique element of T_t with the property that for every $A \in T_t$, $A_1(t) = A - z$ for some $z \ge 0$ and $A_2(t)$ be the unique element of T_t such that for $A \in T_t$, $A_2(t) = A + z$ for some $z \ge 0$. For $\tau \in (0, 1/2)$ and $\eta > 0$ write

$$\begin{aligned}
\bar{\psi}(\tau, \eta) &= \max\bigl[\sup\{P(A_1(t) \vartriangle (A_1(t) - z)): \tau \le t \le 1 - \tau, 0 \le z \le \eta\}, \\
&\qquad \sup\{P(A_2(t) \vartriangle (A_2(t) + z)): \tau \le t \le 1 - \tau, 0 \le z \le \eta\}\bigr], \\
\delta(\tau, \eta) &= \min\bigl[\inf\{P(A_1(t)) - P(A_1(t) - z): \tau \le t \le 1 - \tau, z > \eta\}, \\
&\qquad \inf\{P(A_2(t)) - P(A_2(t) + z): \tau \le t \le 1 - \tau, z > \eta\}\bigr], \\
\varepsilon(\tau, \eta) &= 4\tau + \bar{\psi}(\tau, \eta) + \delta(\tau, \eta).
\end{aligned}$$

With these notations the reader can now easily check the validity of (C_8).

The message of this example is that the asymptotic distribution of the length of the shortt can possibly be nonnormal. In fact it may happen that $B(t)$ is normal for certain values of t and nonnormal for others. When F is the uniform (α, β) distribution, $B(t)$ is nonnormal for each $0 < t < 1$. See the next example.

EXAMPLE 3 (Multivariate uniform). Let $X_1, \dots, X_n$ be i.i.d. uniform I^d, where $I = (0, 1)$ and λ be λ_d.

EXAMPLE 3(a). Set $\mathbb{A} = \{\prod_{i=1}^{d}(0, b_i]: 0 < b_i < 1, i = 1, \dots, d\}$. Here $U(t) = t$ and $T_t = \{A \in \mathbb{A}: \lambda_d(A) = t\}$. All the assumptions (C_1)–(C_8) are satisfied and also (1.9). [Actually, $\lim_{t \uparrow 1} h(t) = 1$ here; hence (1.9) is not needed.] We get from Theorem 1.2 that as $n \to \infty$,

$$\sup_{0<t<1} \left| n^{1/2}(U_n(t) - t) + B(t) \right| \to_{\mathbb{P}} 0.$$

In this case when $d \geq 2$, B is a continuous non-Gaussian process on $[0, 1]$ with $EB(t) > 0$ for all $0 < t < 1$.

EXAMPLE 3(b). Now set $\mathbb{A} = \{\prod_{i=1}^{d}[a_i, b_i]: 0 < a_i < b_i < 1, i = 1, \dots, d\}$. Then everything holds true verbatim as in Example 3(a). Note that the process B that arises in this example is not the same as that of Example 3(a) and is non-Gaussian for all $d \geq 1$.

EXAMPLE 3(c). Everything also works if $\mathbb{A}$ consists of all closed convex subsets of I^d, $d \leq 2$, but not for $d > 2$. Assumption (C_3) no longer holds when $d > 2$; see Bolthausen (1978).

3. Proofs of the theorems. For the proofs of Theorems 1.1 and 1.2 we shall need Proposition 3.1. Before we can state it we need some notation. Set

$$\overline{P}_n(t) = \sup_{\substack{\lambda(A) \leq U(t) \\ A \in \mathbb{A}}} P_n(A), \qquad 0 < t < 1, \tag{3.1}$$

$\overline{P}_n(0) := 0$ and $\overline{P}_n(1) := 1$. Consider the process

$$\overline{\alpha}_n(t) = n^{1/2}\big(\overline{P}_n(t) - t\big), \qquad 0 \leq t \leq 1. \tag{3.2}$$

PROPOSITION 3.1. *Under assumptions* (C_1)–(C_8), *on the probability space of* (1.3),

$$\sup_{0 \leq t \leq 1} \left| \overline{\alpha}_n(t) - B(t) \right| \to 0 \quad a.s.\ as\ n \to \infty. \tag{3.3}$$

PROOF. Notice that for any $0 < t < 1$,

$$\begin{aligned} B(t) - \overline{\alpha}_n(t) &\leq B(t) - n^{1/2}\left\{ \sup_{\substack{\lambda(A) \leq U(t) \\ P(A) = t}} P_n(A) - t \right\} \\ &\leq \sup_{A \in \mathbb{A}} \left| B_P(A) - \alpha_n(A) \right|. \end{aligned} \tag{3.4}$$

Thus by (1.3)

$$\limsup_{n \to \infty} \sup_{0 \leq t \leq 1} \big(B(t) - \overline{\alpha}_n(t)\big) \leq 0 \quad \text{a.s.} \tag{3.5}$$

Also we have for any $0 < t < 1$,

$$\bar{\alpha}_n(t) - B(t) \le \left\{ n^{1/2} \left(\sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} P_n(A) - t \right) - B(t) \right\}$$
$$\vee \left\{ n^{1/2} \left(\sup_{P(A) \le t - n^{-1/4}} P_n(A) - t \right) - B(t) \right\}. \tag{3.6}$$

The second term on the right side of (3.6) is

$$\le n^{1/2} \left(\sup_{P(A) \le t} (P_n(A) - P(A)) \right) + |B(t)| - n^{1/4}.$$

This last term is, uniformly in t,

$$\le 2 \sup_{A \in \mathbb{A}} |B_P(A)| + \sup_{A \in \mathbb{A}} |\alpha_n(A) - B_P(A)| - n^{1/4},$$

which by (1.3) and boundedness of B_P converges to $-\infty$ with probability 1 as $n \to \infty$. Next consider the first term on the right side of (3.6). For any $0 < t < 1$, this term is

$$\le n^{1/2} \sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} (P_n(A) - P(A)) - B(t)$$

$$\le \sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} |\alpha_n(A) - B_P(A)| + \left\{ \sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} B_P(A) - B(t) \right\}.$$

We see now from (1.3) and (3.5) that the proof of (3.3) will be complete if we show that for every $\omega \in \Omega$,

$$\sup_{0 < t < 1} \left\{ \sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} B_P(A) - B(t) \right\} \to 0 \quad \text{as } n \to \infty. \tag{3.7}$$

By (C_8) combined with (C_6) and (C_7), and uniform continuity of B_P for any $\eta > 0$, we have for all large n

$$\sup_{0 < t < 1} \left| \sup_{\substack{\lambda(A) \le U(t) \\ t - n^{-1/4} < P(A) \le t}} B_P(A) - B(t) \right| \le \eta. \tag{3.8}$$

Since $\eta > 0$ is arbitrary, this implies (3.7). □

We introduce notation

$$V_n(s) = \inf\{t \colon \bar{P}_n(t) \ge s, 0 \le t \le 1\}, \qquad 0 \le s \le 1, \tag{3.9}$$

and

$$\bar{\beta}_n(t) = n^{1/2}(V_n(t) - t), \qquad 0 \le t \le 1. \tag{3.10}$$

The following two corollaries are essential to our proofs of Theorems 1.1 and 1.2. For the proof of the first one we need the following version of Lemma 1 in Vervaat (1972).

FACT 3.1. For each $n \geq 1$, let x_n be a nondecreasing function on $[0, 1]$ with $x_n(0) = 0$ and $x_n(1) = 1$. Moreover, let y be a continuous function on $[0, 1]$. If as $n \to \infty$,

$$\sup_{0 \leq t \leq 1} \left| n^{1/2}(x_n(t) - t) - y(t) \right| \to 0,$$

then

$$\sup_{0 \leq t \leq 1} \left| n^{1/2}\left(x_n^{-1}(t) - t\right) + y(t) \right| \to 0,$$

where $x_n^{-1}(t) = \inf\{u: x_n(u) \geq t\}$, $0 \leq t \leq 1$.

COROLLARY 3.1. *On the probability space of* (1.3),

$$\sup_{0 \leq t \leq 1} \left| \bar{\beta}_n(t) + B(t) \right| \to 0 \quad a.s. \text{ as } n \to \infty. \tag{3.11}$$

COROLLARY 3.2. *On the probability space of* (1.3),

$$\sup_{0 \leq t \leq 1} \left| V_n(t) - t \right| \to 0 \quad a.s. \text{ as } n \to \infty. \tag{3.12}$$

Corollary 3.1 is a consequence of Proposition 3.1 and Fact 3.1; Corollary 3.2 follows (obviously) from Corollary 3.1.

PROOF OF THEOREM 1.1. We begin with a lemma. Define $U(0) = L$.

LEMMA 3.1. *With probability* 1 *for all* $0 < s < 1$,

$$U_n(s) = U(V_n(s)). \tag{3.13}$$

PROOF. From the definition of V_n it follows that with probability 1 for all $0 < s < 1$,

$$\begin{aligned} U(V_n(s)) &= \inf\left\{U(t): \sup_{\lambda(A) \leq U(t)} P_n(A) \geq s, 0 < t < 1\right\} \\ &= \inf\left\{r: \sup_{\lambda(A) \leq r} P_n(A) \geq s, L < r < R\right\}. \end{aligned}$$

Recall that

$$U_n(s) = \inf\{\lambda(A): P_n(A) \geq s,\ A \in \mathbb{A}\}.$$

Write

$$S_1 = \left\{r: \sup_{\lambda(A) \leq t} P_n(A) \geq s,\ L < r < R\right\}$$

and

$$S_2 = \{\lambda(A): P_n(A) \geq s\}.$$

If $r \in S_1$, then there exists an $A \in \mathbb{A}$ with $\lambda(A) \le r$ and $P_n(A) \ge s$. Hence there is an $x \le r$ with $x \in S_2$. This implies

$$U_n(s) \le U(V_n(s)) \quad \text{a.s.} \tag{3.14}$$

It remains to show

$$U(V_n(s)) \le U_n(s) \quad \text{a.s.} \tag{3.15}$$

If $r \in S_2$ with $L < r < R$, then there exists an $A \in \mathbb{A}$ with $\lambda(A) = r$ and $P_n(A) \ge s$. Hence

$$\sup_{\lambda(A) \le r} P_n(A) \ge s,$$

which yields $r \in S_1$. This implies (3.15). □

We are now ready to complete the proof of Theorem 1.1. For each $a \le t \le b$ we get by the mean value theorem and Lemma 3.1, that almost surely

$$\beta_n(t) + B(t) = \frac{g(t)}{g(\theta_n)}\{\bar{\beta}_n(t) + B(t)\} - \left(\frac{g(t)}{g(\theta_n)} - 1\right)B(t), \tag{3.16}$$

where θ_n lies between t and $V_n(t)$. Assertion (1.8) is now a trivial consequence of (3.11) and (3.12). □

Proof of Theorem 1.2. The proof of Theorem 1.2 requires three additional lemmas.

Lemma 3.2. *Let $\{Y_{n,k}\}_{n \ge 1, k \ge 1}$ be a double sequence of random variables such that for each $n, k \in \mathbb{N}$, $Y_{n,k}$ is binomial $(n, 2^{-k})$. Then*

$$Y_n := \sup_{k \in \mathbb{N}} 2^k Y_{n,k}/n = O_{\tilde{\mathbb{P}}}(1) \quad \text{as } n \to \infty, \tag{3.17}$$

where $\tilde{\mathbb{P}}$ denotes the probability measure on the space on which these random variables are defined.

Proof. Choose any $\varepsilon > 0$. It suffices to find a $D > 0$ such that

$$\tilde{\mathbb{P}}(Y_n \ge D) \le \varepsilon \quad \text{for all } n. \tag{3.18}$$

We have for any $D > 0$,

$$\tilde{\mathbb{P}}(Y_n \ge D) \le \sum_{k=1}^{\infty} \tilde{\mathbb{P}}(Y_{n,k} \ge Dn2^{-k}).$$

Notice that if $Dn2^{-k} < 1$, then

$$\tilde{\mathbb{P}}(Y_{n,k} \ge Dn2^{-k}) = \tilde{\mathbb{P}}(Y_{n,k} > 0) = 1 - (1 - 2^{-k})^n \le n2^{-k}.$$

Hence

$$\sum_{k:\, Dn2^{-k} < 1} \tilde{\mathbb{P}}(Y_{n,k} \ge Dn2^{-k}) \le \sum_{k:\, Dn2^{-k} < 1} n2^{-k} \le 2/D. \tag{3.19}$$

Now consider $Dn2^{-k} \geq 1$. In this case for D large and with $\tilde{D} = D - 1$,

$$\tilde{\mathbb{P}}\big(Y_{n,k} \geq Dn2^{-k}\big) = \tilde{\mathbb{P}}\big(n^{-1/2}(Y_{n,k} - n2^{-k}) \geq \tilde{D}n^{1/2}2^{-k}\big),$$

which by Bennett's inequality [see, e.g., Shorack and Wellner (1986), page 440] is

$$\leq \exp\big(-\tilde{D}^2 n 2^{-k-1}\psi(\tilde{D})\big),$$

where ψ is a function that $\psi(z) \sim (2\log z)/z$ as $z \to \infty$. Thus for all large enough $\tilde{D}$,

$$\sum_{k:\, Dn2^{-k} \geq 1} \tilde{\mathbb{P}}\big(Y_{n,k} \geq Dn2^{-k}\big) \leq \sum_{i \in \mathbb{N}} \exp\big(-2^{i-3}\log \tilde{D}\big) \leq 2\tilde{D}^{-1/4}. \tag{3.20}$$

Inequalities (3.19) and (3.20) yield (3.18). □

LEMMA 3.3. *On the probability space of* (1.3),

$$\sup_{0<t<1} (1-t)/(1-V_n(t)) = O_{\mathbb{P}}(1). \tag{3.21}$$

PROOF. For any $k \in \mathbb{N}$ choose $A_k \in T_{1-2^{-k}}$ and for $1 - 2^{-k} \leq t < 1 - 2^{-k-1}$ set $A_t = A_k$. Now we have

$$\begin{aligned}\sup_{0<t<1} (1-t)/(1-V_n(t)) &= \sup_{0<t<1} \big(1-\overline{P}_n(t)\big)/(1-t)\\ &= \sup_{0<t<1} \inf_{\lambda(A)\leq U(t)} (1-P_n(A))/(1-t)\\ &\leq 2 \vee \Big\{\sup_{1/2\leq t<1} (1-P_n(A_t))/(1-t)\Big\}\\ &\leq 2 \vee \Big\{\sup_{k\geq 1} 2^{k+1}(1-P_n(A_k))\Big\}.\end{aligned}$$

An application of Lemma 3.2 now completes the proof. □

LEMMA 3.4. *Under the assumptions of Theorem* 1.2, *especially* (1.9), *there exists a* $\delta > 0$ *such that for* $1 - \delta \leq t \leq s < 1$,

$$g(t)/g(s) \leq ((1-t)/(1-s))^M. \tag{3.22}$$

PROOF. Simple analysis; see, for example, Shorack and Wellner [(1986), page 644]. □

We are now prepared to finish the proof of Theorem 1.2. An easy adaptation of the proof of Theorem 1.1 using the assumption that $0 < h(L+) < \infty$ shows that for all $0 < \eta < 1$,

$$\sup_{0<t\leq 1-\eta} |\beta_n(t) + B(t)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.23}$$

If $\lim_{t\uparrow R} h(t) > 0$, necessarily meaning $R < \infty$, then (3.23) holds with $\eta = 0$ by the same proof as before. So from now on we assume $\lim_{t\uparrow R} h(t) = 0$. From

(3.16) we obtain that for any $0 < \eta < 1$ almost surely

$$\sup_{1-\eta \le t < 1} |\beta_n(t) + B(t)| \le \sup_{1-\eta \le t < 1} \left| \frac{g(t)}{g(\theta_n)} (\bar{\beta}_n(t) + B(t)) \right| + \sup_{1-\eta \le t < 1} \left| \left(\frac{g(t)}{g(\theta_n)} - 1 \right) B(t) \right|.$$

Next by Corollary 3.2, with probability 1 for all small enough $\eta > 0$,

$$\sup_{1-\eta \le t < 1} g(t)/g(\theta_n) \le \sup_{1-\eta \le t < 1} g(t)/g(t \vee \theta_n) \tag{3.24}$$

for all large enough n, which by Lemma 3.4 is

$$\le \sup_{1-\eta \le t < 1} \left(\frac{1-t}{(1-t) \wedge (1 - V_n(t))} \right)^M.$$

This last expression is, in turn by Lemma 3.3, $O_{\mathbb{P}}(1)$. Thus we obtained that for all small enough $\eta > 0$,

$$\sup_{1-\eta \le t < 1} g(t)/g(t \vee \theta_n) = O_{\mathbb{P}}(1). \tag{3.25}$$

This in combination with Corollary 3.1 says that for all small $\eta > 0$,

$$\sup_{1-\eta \le t < 1} \left| \frac{g(t)}{g(\theta_n)} (\bar{\beta}_n(t) + B(t)) \right| = o_{\mathbb{P}}(1). \tag{3.26}$$

Now for all $\eta > 0$ small, with probability 1 for all large n,

$$\begin{aligned} &\sup_{1-\eta \le t < 1} \left| \left(\frac{g(t)}{g(\theta_n)} - 1 \right) B(t) \right| \\ &\quad \le 2 \sup_{1-\eta \le t < 1} \frac{g(t)}{g(t \vee \theta_n)} \sup_{1-\eta \le t < 1} |B(t)|. \end{aligned} \tag{3.27}$$

From continuity of B at 1 and $B(1) = 0$ a.s. we have

$$\lim_{\eta \downarrow 0} \sup_{1-\eta \le t < 1} |B(t)| = 0 \quad \text{a.s.} \tag{3.28}$$

Finally a routine argument based on (3.23), (3.26), (3.27), (3.25) and (3.28) establishes that (1.10) holds. □

Acknowledgments. We thank the referees for careful readings of the manuscript and a number of useful suggestions. In particular, the term *shortt* was coined by one of the referees.

REFERENCES

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

Barnett, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A* **139** 318–344.

Bolthausen, E. (1978). Weak convergence of an empirical process indexed by the closed subsets of I^2. *Z. Wahrsch. Verw. Gebiete* **43** 173–181.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications. Regional Conference Series on Applied Mathematics* **42**. SIAM, Philadelphia.

CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882–894.

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DEHEUVELS, P., EINMAHL, J. H. J., MASON, D. M. and RUYMGAART, F. H. (1988). The almost sure behavior of maximal and minimal multivariate k_n-spacings. *J. Multivariate Anal.* **24** 155–176.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 141–178. Springer, Berlin.

EDDY, W. F. (1985). Ordering of multivariate data. In *Computer Science and Statistics: Proceedings of the 16th Symposium on the Interface* (L. Billard, ed.) 25–30. North-Holland, Amsterdam.

GAENSSLER, P. (1983). *Empirical Processes.* IMS, Hayward, Calif.

GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

PLACKETT, R. L. (1976). Discussion of "The ordering of multivariate data," by V. Barnett. *J. Roy. Statist. Soc. Ser. A* **139** 344–346.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.

PYKE, R. (1975). Multidimensional empirical processes: Some comments. In *Stochastic Processes and Related Topics, Proceedings of the Summer Research Institute on Statistical Inference for Stochastic Processes* **2** 45–58. Academic, New York.

PYKE, R. (1984). Comment on "Some limit theorems for empirical processes" by E. Giné and J. Zinn. *Ann. Probab.* **12** 996–997.

PYKE, R. (1985). Opportunities for set-indexed empirical and quantile processes in inference. In *Bulletin of the International Statistical Institute: Proceedings of the 45th Session, Invited Papers* **51** 25.2.1–25.2.12. ISI, Voozburg, The Netherlands.

REISS, R.-D. (1989). *Approximate Distribution of Order Statistics.* Springer, New York.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

ROUSSEEUW, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics with Applications* (W. Grossman, G. Pflug, I. Vincze, W. Wertz, eds.) 283–297. Reidel, Dordrecht.

ROUSSEEUW, P. and LEROY, A. (1988). A robust scale estimator based on the shortest half. *Statist. Neerlandica* **42** 103–116.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (R. D. James, ed.) **2** 523–531. Canadian Math. Congress.

VERVAAT, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Gebiete* **23** 245–253.

DEPARTMENT OF MATHEMATICS
AND COMPUTING SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P. O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS

DEPARTMENT OF MATHEMATICAL SCIENCES
501 EWING HALL
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19716