ELSEVIER Stochastic Processes and their Applications 69 (1997) 1–24

stochastic processes and their applications

# Minimum volume sets and generalized quantile processes

Wolfgang Polonik*

*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120, Heidelberg, Germany*

Received 4 April 1995; received in revised form 6 February 1997

**Abstract**

Bahadur–Kiefer approximations for generalized quantile processes as defined in Einmahl and Mason (1992) are given which generalize results for the classical one-dimensional quantile processes. An as application we consider the special case of the volume process of minimum volume sets in classes $\mathbb{C}$ of subsets of the $d$-dimensional Euclidean space. Minimum volume sets can be used as estimators of level sets of a density and might be useful in cluster analysis. The volume of minimum volume sets itself can be used for robust estimation of scale. Consistency results and rates of convergence for minimum volume sets are given. Rates of convergence of minimum volume sets can be used to obtain Bahadur–Kiefer approximations for the corresponding volume process and vice versa. A generalization of the minimum volume approach to non-i.i.d. problems like regression and spectral analysis of time series is discussed. © 1997 Elsevier Science B.V.

*Keywords*: Bahadur–Kiefer approximation; Empirical process theory; Generalized uniform empirical process; Level set estimation; Rates of convergence; Test for multimodality

## 1. Introduction

The asymptotic behaviour of minimum-volume-sets (MV-sets) and of the generalized quantile process as defined in Einmahl and Mason (1992) is studied in this paper. We give consistency results and rates of convergence of MV-sets and Bahadur–Kiefer approximations for the generalized quantile process. The results show that rates for MV-sets can be used to obtain rates for the generalized quantile process and vice versa. Empirical process theory is the main mathematical tool.

The setup is as follows. Let $X_1, X_2, \dots, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$ with distribution $F$. Further , let $\mathbb{C}$ be a class of measurable of $\mathbb{R}^d$ and let $\lambda$ denote a real-valued function defined on $\mathbb{C}$. Define the quantile function based on $F$, $\lambda$ and $\mathbb{C}$ as

$$V(\alpha) = \inf\{\lambda(C)\colon F(C) \geqslant \alpha,\ C \in \mathbb{C}\}, \quad 0 < \alpha < 1 \tag{1.1}$$

* E-mail: polonik@statlab.uni-heidelberg.de.


*PII* S0304-4149(97)00028-8

The *empirical* quantile function is defined as

$$V_n(\alpha) = \inf\{\lambda(C)\colon F_n(C) \geqslant \alpha,\ C \in \mathbb{C}\}, \quad 0 < \alpha < 1, \tag{1.2}$$

where $F_n$ denotes the empirical distribution of the first $n$ observations, which puts mass $n^{-1}$ at each of the observations $X_1, X_2, \dots, X_n$. With $\mathbb{C} = \{(-\infty, t), t \in \mathbb{R}\}$ and $\lambda((-\infty, t]) = t$ the quantities $V(\alpha)$ and $V_n(\alpha)$ are the classical real-valued $\alpha$-quantile and empirical $\alpha$-quantile, respectively. Suppose that $V$ is differentiable with derivative $v$, then

$$q_n(\alpha) = (v(\alpha))^{-1} n^{1/2}(V_n(\alpha) - V(\alpha)) \tag{1.3}$$

is the standardized *generalized quantile process* as defined by Einmahl and Mason (1992). The factor $(v(\alpha))^{-1}$ is the analogue to the well-known factor $f(F^{-1}(\alpha))$ of the classical one-dimensional standardized quantile process. In case of existence we denote a minimizing set in the definition of $V(\alpha)$ and $V_n(\alpha)$ by $Q(\alpha)$ and $Q_n(\alpha)$, respectively, i.e. we have

$$V(\alpha) = \lambda(Q(\alpha)) \quad \text{and} \quad V_n(\alpha) = \lambda(Q_n(\alpha)).$$

For $\lambda(C) = \mathrm{Leb}(C)$, where Leb denotes Lebesgue measure, these minimizing sets are called *minimum-volume* (*MV*)*-sets in* $\mathbb{C}$ (*with respect to F or* $F_n$, respectively). We refer to this situation as the MV-case and sometimes write $\tilde{q}_n$, $\tilde{V}_n$, $\tilde{V}$ instead of $q_n$, $V_n$, $V$, respectively, to distinguish between the MV-case and the general case. $\tilde{q}_n$ is called *volume process*. In the present paper we derive weak Bahadur–Kiefer approximations for $q_n(\alpha)$, i.e. we derive stochastic rates of convergence for $\sup_{\eta<\alpha<1-\eta}|q_n(\alpha) + \nu_n(Q(\alpha))|$, where $\nu_n(C) = n^{1/2}(F_n - F)(C)$ is the $\mathbb{C}$-indexed empirical process and $\eta \geqslant 0$. We also study the asymptotic behaviour of $Q_n(\alpha)$.

The MV-case is an important special case of the presented approach. MV-sets have been studied in the context of robust statistics. In the one-dimensional case Andrews et al. (1972) used the mean of all data points inside a MV-interval as a robust estimator of location which they called "shorth", or "$\alpha$-shorth". Nowadays, in the literature often MV-intervals itselves are called "shorth" or "$\alpha$-shorth". (Even earlier than Andrews et al., Lientz (1970) investigated a certain localized approach. For every fixed $x \in \mathbb{R}$, he used MV-intervals in the class of all intervals which contain $x$. See also Sawitzki (1994)). In higher dimensions, Sager (1978, 1979) considered MV-sets in classes of polynomial regions and in the class of convex sets in $\mathbb{R}^2$. Rousseeuw (1986) used MV-ellipsoids to construct robust estimators of location and dispersion parameters. Davies (1987) studied these estimators in the context of $S$-estimators. The volume of MV-sets, $\tilde{V}_n(\alpha)$, can be used for scale estimation. This has first been considered in Grübel (1988) in the onedimensional case. There $\tilde{V}_n(\alpha)$ is the length of the $\alpha$-shorth. In our notation Grübel proved that under certain smoothness assumptions $\tilde{q}_n(\alpha)$ with $\mathbb{C}$ the class of closed intervals converges weakly to a Brownian Bridge if $\alpha$ is bounded away from zero and one. Einmahl and Mason (1992) generalized this result in proving that under certain conditions $\sup_{0<\alpha<1}|q_n(\alpha) + B_n(\alpha)|$ converges to zero in probability as $n$ tends to infinity, where $B_n$ are versions of standard Brownian bridges. They also proved almost sure convergence for $\alpha$ bounded away from 0 to 1. Note that if $f(x) = f_0((x-\mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$, then the lengths of the level sets of

$f$ equal $\sigma$ times the lengths of the corresponding level sets of $f_0$. As already mentioned in Einmahl and Mason this property can be used to generalize the well known QQ-plots by plotting $\tilde{V}_0(\alpha) = \mathrm{Leb}(Q_0(\alpha))$ against $\tilde{V}_n(\alpha) = \mathrm{Leb}(Q_n(\alpha))$, where $Q_0(\alpha)$ denotes the MV-set with respect to $f_0$.

$\tilde{V}_n(\alpha)$ can also be used to investigate modality of a distribution (cf. Section 5). If for example the underlying univariate distribution is bimodal then there exists an $\alpha > 0$ such that $\tilde{V}(\alpha)$ in the class of all unions of two intervals is smaller than $\tilde{V}(\alpha)$ in the class of intervals. Therefore the (scaled) difference of the corresponding empirical versions can be used for testing unimodality. This idea is related to the idea of using excess mass estimates to investigate the modality of a distribution proposed by Müller and Sawitzki (1987) and Hartigan (1987) (for generalizations see Nolan, 1989, and Polonik, 1995b).

Another important statistical aspect of the MV-approach is level set estimation. Let $f$ denote the Lebesgue density of $F$ and let $\Gamma(\mu) = \{f(x) \geqslant \mu\}$ denote a level set of $f$. If $\Gamma(\mu) \in \mathbb{C}$ then $\tilde{V}(\alpha) = \mathrm{Leb}(\Gamma(\mu_\alpha))$, where $\mu_\alpha > 0$ is chosen such that $F(\Gamma(\mu_\alpha) = \alpha$ (cf. Fig. 1). Therefore a natural level set estimator is given by the empirical counterparts $Q_n(\alpha)$. Level set estimation is useful especially in cluster analysis, where one is interested in regions which contain high mass concentration. (The case $\alpha = 1$ corresponds to estimation of the support of the underlying distribution.) For recent work in the area of level sets estimation see for example Cuevas (1990), Cuevas and Fraiman (1993) for support estimation, Molchanov (1993) for estimating level sets by means of density estimation and Tsybakov (1997) for minimax rates of convergence for level set estimators.

Let us briefly discuss the choice of $\mathbb{C}$. First note, that richness of $\mathbb{C}$ (measured by metric entropy) influences asymptotic properties of estimators and tests as considered in this paper. The richer $\mathbb{C}$ the slower are the rates of convergence. Richness of $\mathbb{C}$ is also crucial for time needed for calculation of the procedures proposed in this paper. From this point of view rich classes are worse than sparse classes. In the MV-case a further aspect comes in. There the assumption that all the level sets $\Gamma(\mu)$ of $f$ lie in $\mathbb{C}$ is crucial. Through this assumption richness of $\mathbb{C}$ means richness of the statistical model. Note that by appropriate choice of $\mathbb{C}$ it is possible to model quantitative aspects of the underlying density such as shape of level sets (e.g. convexity), symmetry, monotonicity, modality (see Polonik, 1995a, 1995b). For example in the one dimensional case monotone decreasing densities $[0, \infty)$ can be modeled by choosing $\mathbb{C} = \{[0, x], x \in \mathbb{R}\}$. Summing up it may be said, that for an appropriate choice of $\mathbb{C}$ one has to balance between statistical properties, time needed for calculation and richness of the model. In this paper we do not specify a class $\mathbb{C}$. We consider general types of classes such as Glivenko Cantelli classes, Vapnik Cervonenkis classes, Donsker classes or more general classes which satisfy certain entropy conditions. The classes of invervals, ellipsoids and convex sets are special cases.

The present paper is organized as follows. First the asymptotic behaviour of $Q_n(\alpha)$ is studied. Consistency of $Q_n(\alpha)$ as an estimator of $Q(\alpha)$ is shown in Section 2 and rates of convergence are given in Section 3 for case $\lambda = \mathrm{Leb}$. As a (pseudo) metric on $\mathbb{C}$ the $F$-measure of symmetric difference is used. In Section 4 Bahadur–Kiefer approximations of the generalized quantile process are given, where the results are sharper for the case $v =$ Leb. In Section 5 we study tests for multimodality based on the volume of MV-sets. Section 6 contains some extensions and generalizations. It is indicated that the MV-approach can also be applied to regression problems and to spectral analysis in the time series context, and that it can be used to handle processes which appear in the context of multivariate trimming in Nolan (1992). Section 7 contains all the proofs.

## 2. Consistency results

Let $(\Omega, P)$ denote the underlying probability space. In order to avoid measurability considerations we define for a function $f: \Omega \to \mathbb{R}$ the *measurable cover function* $f^*$ as the smallest measurable function from $\Omega$ to $\mathbb{R}$ lying everywhere above $f$ (see e.g. Dudley, 1984). Furthermore, let $P^*$ and $P_*$ denote *outer and inner probability*, respectively.

**Definition.** A class $\mathbb{C}$ of measurable subsets of $\mathbb{R}^d$ is called a *Glivenko Cantelli* (GC)-*class* for $F$, iff

$$\|F_n - F\|_{\mathbb{C}}^* := (\sup_{C \in \mathbb{C}} |F_n(C) - F(C)|)^* \to 0 \quad \text{a.s.}$$

In what follows we denote some main assumptions by (A0), (A1), … , etc.

(A0) $\mathbb{C}$ and $\lambda$ are such that $(V_n(\alpha))^* < \infty$ a.s. for all $\alpha \in [0, 1]$.
(A1) $\mathbb{C}$ is a GC-class for $F$.
The following proposition will be used below to derive consistency of $Q_n(\alpha)$.

**Proposition 2.1.** *Suppose that* (A0) *and* (A1) *hold. If* $V(\cdot)$ *is continuous in* $\alpha$, *then*

$$|V_n(\alpha) - V(\alpha)|^* \to 0 \quad \text{a.s.}$$

*The convergence is uniform in* $\alpha \in A$ *if* $U(\cdot)$ *is uniformly continuous in* $A \subset [0, 1]$.

In the MV-case continuity of $V$ holds in following situations:

(i) Suppose that $F$ has a bounded Lebesgue density $f$ without flat parts, i.e. $F\{x: f(x) = \mu\} = 0$ of all $\mu$. If the level sets $\Gamma(\mu)$ (for $\mu = 0$ we define $\Gamma(0)$ to be the support of $F$) all lie in $\mathbb{C}$, then $V$ is continuous in (0, 1) and uniformly continuous in $(0, 1 - \varepsilon]$ for every $\varepsilon > 0$. If the support of $f$ is bounded then $V$ is uniformly continous in [0, 1].

(ii) Let $f$ be a density on the real line which is bounded and unimodal in the sense that there exists a point $x_0$ such that $f$ is non-decreasing to the left of $x_0$ and non-increasing to the right. Choose $\mathbb{C}$ as the class of all intervals. Then $V$ is continuous. This is easy to see, because if $V$ would be discontinuous at some $\alpha < 1$, then the inverse function would have a flat part. But that would mean that the maximal probability content of an interval of given length could not be increased by increasing the length, which would give a contradiction.

In order to formulate the next proposition we need some further assumptions:

(A2) *$F$ has a bounded Lebesgue density $f$.*

(A3) *For every $\alpha \in [0, 1]$ there exists a unique (up to $F$-nullsets) set $Q(\alpha)$ with $F$-measure $\alpha$*

(A4) *For every* $\alpha \in [0, 1]$ *there exists a set* $Q_n(\alpha)$.
(A5) $\lambda$ *is lower semicontinuous for* $d_F$.

Assumption (A3) says that $\mathbb{C}$ has to be rich enough. Note that by definition $Q(\alpha)$ has $F$-measure $\geqslant \alpha$. Here we require $Q(\alpha)$ to have $F$-measure exactly equal to $\alpha$. (A3) holds in the situations (i) and (ii) given above. Whereas $Q(\alpha)$ is assumed to be essentially unique, (A3), this is not required for the empirical sets $Q_n(\alpha)$. *The results given below hold for every choice of* $Q_n(\alpha)$. As a pseudo metric on the class $\mathbb{C}$ we use

$$d_F(C, D) := F(C \Delta D), \quad C, D \in \mathbb{C}.$$

**Proposition 2.2.** *Suppose that* (A0)–(A5) *hold. Let* $\eta \geqslant 0$. *Assume that*

(i) *there exists a distribution G with positive Lebesgue density such that* $(\mathbb{C}, d_G)$ *is quasi compact*
(ii) $V$ *is continuous in* $[\eta, 1-\eta]$

*then*

$$\left( \sup_{\eta \leqslant \alpha \leqslant 1-\eta} d_F(Q_n(\alpha), Q(\alpha)) \right)^* \to 0 \quad a.s.$$

**Remark.** In special cases (i.e. for special classes $\mathbb{C}$) consistency of $Q_n(\alpha)$ has been proven in the literature before. Consistency of classical quantiles, and consistency of MV-ellipsoids are well known. There of course one uses the Euclidean distance. In non-parametric cases consistency of convex MV-sets (in Hausdorff-distance) is proven in Sager (1979).

## 3. Rates of convergence of $Q_n(\alpha)$

In this section we only consider the MV-case, i.e. the case $\lambda = \text{Leb}$. To formulate the results in this section we need to introduce some additional terminology and notation. For the proofs of the theorems given below we shall use results of Alexander (1984, 1985) about the behaviour of the empirical process. For that reason we also use some of his terminology. Alexander considers empirical processes indexed by VC-classes which he called "$n$-deviation measurable". Here we do not give the definition of "$n$-deviation measurable", because all the standard VC-classes which we are interested in (the classes of intervals, balls, ellipsoids in $\mathbb{R}^d$ and finite unions and differences of them) satisfy this measurability condition. Alexander calls a class $\mathbb{C}$ *m-constructible* out of $\mathbb{D}$, if $\mathbb{C}$ can be constructed out of $\mathbb{D}$ by means of $m \in \mathbb{N}$ set theoretic operations, $\cap$, $\cup$ $\setminus$. For $v \in \mathbb{N}$ a class $\mathbb{C}$ is called *$(v, m)$-constructible VC-class* if $\mathbb{C}$ is $m$-constructible from a VC-class $\mathbb{D}$ whose index is smaller than or equal to $v$. The index of a VC-class is defined as the smallest integer $k$, such that $\mathbb{D}$ "shatters" no set which consists of $k$ points. And $\mathbb{D}$ "shatters" a finite set $C$, iff every $B \subset C$ is of the form $C \cap D$ for some $D \in \mathbb{D}$. We also need the notion of *metric entropy with inclusion* of $\mathbb{C}$ with respect to $F$. Let

$$N_I(\varepsilon, \mathbb{C}, F) := \inf\{m \in \mathbb{N}: \exists C_1, \ldots, C_m \textit{ measurable, such that for every } C \in \mathbb{C} \textit{ there exist } i, j \in \{1, \ldots, m\} \textit{ with } C_i \subset C \subset C_j \textit{ and } F(C_j \backslash C_i) < \varepsilon\},$$

*then* $\log N_1(\varepsilon, \mathbb{C}, F)$ is called metric entropy with inclusion of $\mathbb{C}$ with respect to $F$. For a set $A \subset \mathbb{R}^d$ and $\varepsilon > 0$ we denote with $A^\varepsilon$ the set $A$ blown up by $\varepsilon$, i.e. the set which consists of the union of all closed $\varepsilon$-balls around points in $A$.

As briefly noted in the introduction, metric entropy measures richness (or dimensionality) of the class $\mathbb{C}$, and it is used in empirical process theory to control the asymptotic behaviour of the $\mathbb{C}$-indexed empirical process. The same is true for the VC-property. We shall assume roughly (cf. Theorem 3.1, Part Ib) that metric entropy behaves like a polynome in $\varepsilon > 0$ of degree $r > 0$. Separately we consider the case that $\mathbb{C}$ is a VC-class. It can be shown that for VC-classes metric entropy behaves like $O(\log \varepsilon)$ for $\varepsilon \to 0$. This fact is reflected in the rates of convergence given below, namely, up to a log-term the below given rates of convergence for VC-classes $\mathbb{C}$ can be obtained from the rates given under metric entropy conditions by formally replacing $r$ through 0.

We need some additional assumptions:

(A6) The sets $\Gamma(\mu) = \{x \in \mathbb{R}^d: f(x) \geqslant \mu\}$, $\mu \geqslant 0$ all lie in $\mathbb{C}$.
(A7) $\mathbb{C}$ is such that $\sup_{0 \leqslant \alpha \leqslant 1} |F_n(Q_n(\alpha)) - \alpha|^* = O(1/n)$ a.s.

Assumption (A7) is satisfied for standard classes like closed intervals (in the univariate case) circles, ellipsoids or convex sets (for higher dimensions) and the corresponding $m$-constructible classes.

In the following theorem the quantity $\mu_\alpha$ defined in the introduction (cf. Fig. 1) becomes important. The reason is that in the situation of Theorem 3.1 we have $\tilde{v}(\alpha) = \tilde{V}'(\alpha) = 1/\mu_\alpha$, and this derivative appears in the definition of $\tilde{q}_n(\alpha)$ (see (1.3)).

**Theorem 3.1.** *Let $\lambda = Leb$. Suppose that (A0)–(A7) hold. Let $A \subset [0, 1]$ and suppose that there exist constants $\gamma$, $C \geqslant 0$ such that for all $\eta > 0$ small enough*

$$\sup_{\alpha \in A} F\{x \in \mathbb{R}^d: |f(x) - \mu_\alpha| < \eta\} \leqslant C\eta^\gamma. \tag{3.1}$$

*Part* I: *If in addition*

(i) $\alpha \to \mu_\alpha$ *is Lipschitz continuous in* $\alpha \in A^\varepsilon \cap [0, 1]$ *for some* $\varepsilon > 0$, *and*
(ii) $\inf_{\alpha \in A} \mu_\alpha > 0$
*then we have the following*:
(a) *If* $\mathbb{C}$ *is an n-deviation measurable* $(v, m)$*-constructible VC-class then for* $\delta > \gamma/(2 + \gamma)$

$$\sup_{\alpha \in A} d_F(Q_n(\alpha), Q(\alpha)) = O_{P^*}(n^{-\delta}) \quad \text{as } n \to \infty. \tag{3.2}$$

(b) *If* $\mathbb{C}$ *is such that for some* $A$, $r > 0$

$$\log N_1(\varepsilon, \mathbb{C}, F) \leqslant A\varepsilon^{-r} \quad \forall \varepsilon > 0 \tag{3.3}$$

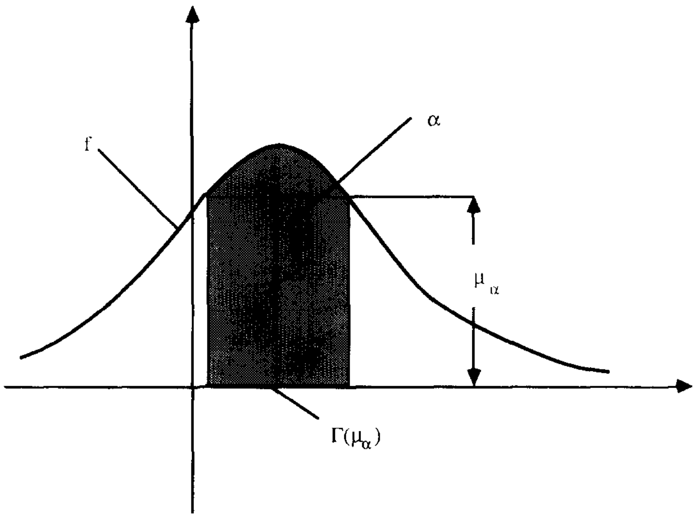


Fig. 1. $\mu_\alpha$ is defined through $F(\Gamma(\mu_\alpha)) = \alpha$

*then as* $n \to \infty$

$$\sup_{\alpha \in A} d_F(Q_n(\alpha), Q(\alpha)) = \mathrm{O}_{\mathrm{p}^*}(\delta_n(r)) \tag{3.4}$$

*where*

$$\delta_n(r) = \begin{cases} n^{-\delta} \text{ for some } \delta > \gamma/(2 + (1+r)\gamma) & \text{if } r < 1, \\ n^{-\gamma/2(1+\gamma)} \log n & \text{if } r = 1, \\ n^{-\gamma/(1+r)(1+\gamma)} & \text{if } r > 1. \end{cases}$$

*Part* II: *Suppose instead of* (i) *and* (ii) *that* $f$ *is continuously differentiable such that*

$$\sup_{0<\alpha<1} \alpha(1-\alpha)\left(\mu_\alpha^2 \int_{\partial\Gamma(\mu_\alpha)} \|grad\, f(s)\|^{-1}\,\mathrm{d}s\right)^{-1} < c < \infty, \tag{3.5}$$

*where* $\partial\Gamma(\mu_\alpha) = \{x: f(x) = \mu_\alpha\}$ *and "ds" denotes the* $(d-1)$*-dimensional Hausdorff measure. Then the assertions of Part* I *hold with* $A = [0, 1]$.

**Remarks.** (i) The same rates of convergence as in Theorem 3.1 also appear in Polonik (1995b) in the context of estimating level sets by means of the so called excess mass approach. There one considers different (but related) level set estimators for which that same upper bounds for the rates of convergence can be shown as for the empirical MV-sets.

(ii) The case $A = [0, 1]$ is formulated separately in Part II, because for this case the conditions of Part I are quite restrictive. In particular, assumption (ii) of Part I is satisfied for $A = [0, 1]$ only if $f$ is bounded away from zero inside a compact support.

(iii) In regular situations we have for fixed $\alpha$, i.e. $A = \{\alpha\}$, that (3.1) holds with $\gamma = 1$. Such levels $\alpha$ will be called *regular*. For regular $\alpha$ Theorem 3.1 gives the following rates. If $\Gamma(\mu_\alpha)$ lies in $\mathbb{C}$ then for $\varepsilon > 0$ arbitrary

$$d_F(Q_n(\alpha), \Gamma(\mu_\alpha)) = \begin{cases} O_{P^*}(n^{-1/3+\varepsilon}) & \text{for } VC \text{ classes } \mathbb{C}, \\ O_{P^*}(n^{-2/7+\varepsilon}) & \text{for } \mathbb{C} = \mathrm{C}^2(K), \\ O_{P^*}(n^{-1/4}\log n) & \text{for } \mathbb{C} = \mathrm{C}^3(K), \\ O_{P^*}(n^{-1/(d+1)}) & \text{for } \mathbb{C} = \mathrm{C}^d(K),\ d > 3, \end{cases}$$

where $\mathrm{C}^d(K)$ denotes the class of convex sets in $\mathbb{R}^d$, $d \geqslant 2$, lying in a compact set $K$. The assertion for the convex sets follows since $r = (d-1)/2$ for $\mathbb{C} = \mathrm{C}^d(K)$ (Dudley, 1984). Tsybakov (1997) shows (under slightly different smoothness assumptions) that $n^{-2/7}$ is the minimax rate for estimating convex level sets.

Since intervals and ellipsoids form VC-classes Theorem 3.1 gives us the almost $n^{-1/3}$ rate for these cases. Note that it is known that the center of the MV-interval (the shorth) and also the center of the MV-ellipsoid converge at an $n^{-1/3}$ rates (cf. Andrews et al., 1972), for the one-dimensional case). Of course one should expect that the $\varepsilon$ can be removed from the rates, but at present we do not know how to do this.

(iv) Let us briefly discuss the validity of assumption (3.5). First note that (3.5) is a special version of the general assumption $\sup_{0<\alpha<1} \alpha(1-\alpha)|v'(\alpha)|(v(\alpha))^{-1} < c < \infty$ which will be used in Section 4 (cf. discussion of assumptions of Theorem 4.2). If we consider only values of $\alpha$ close to one, and restrict the supremum in (3.5) to such $\alpha$, then the tail behaviour of $f$ determines the validity of (3.5). In the one-dimensional case (3.5) holds for example for normal distributions, logistic distributions and exponential distributions (see Shorack and Wellner, 1986, p. 644). If the supremum is (3.5) is restricted to $\alpha$ close to zero and $d = 1$, then for (3.5) to be satisfied we only need $f'$ to be bounded inside $\Gamma(\mu_\alpha)$, for some $\alpha > 0$. However, for $d \geqslant 2$ the integral $\int_{\partial\Gamma(\mu_\alpha)} \|\mathrm{grad} f(s)\|^{-1}\, ds$ can come close to zero even if $\|\mathrm{grad} f(s)\|$ is bounded, because the $(d-1)$-dimensional Hausdorff measure of $\partial\Gamma(\mu_\alpha)$ can become small. For example assume that $f$ has a mode in 0 and that locally around zero $f(x) = -\|x\|^k + c$, for some $c$, $k > 0$. Then it is easy to verify that $\int_{\partial\Gamma(\mu_\alpha)} \|\mathrm{grad} f(s)\|^{-1}\, ds \approx \alpha^{1-k/d}$. Hence the integral $\int_{\partial\Gamma(\mu_\alpha)} \|\mathrm{grad}\, f(s)\|^{-1}\, ds$ converges to zero if $k < d$. However, $\alpha(\int_{\partial\Gamma(\mu_\alpha)} \|\mathrm{grad}\, f(s)\|^{-1}\, ds)^{-1} \approx \alpha^{k/d}$, such that (3.5) is satisfied (for $\alpha$ close to 0) for each $k > 0$.

The main technical result for deriving the rates of Theorem 3.1 is inequality (3.6) which is given in the following lemma. We formulate it here, since it shows that the analysis of $d_F((Q_n(\alpha), \Gamma(\mu_\alpha))$ can be decomposed into a deterministic and a stochastic term. Somehow, this is like a bias-variance decomposition. It also shows, how condition (3.1) and the conditions on the empirical process come in. A similar inequality has been used in Polonik (1995b) in the context of estimating level sets by means of the excess mass approach.

**Lemma 3.2.** *Let* $\lambda = Leb$. *Suppost that* (A0), (A3), (A4), (A6) *hold and let* $M = \sup_{x\in\mathbb{R}} |f(x)| (< \infty$ *by* (A3)). *Let* $\alpha \in (0,1)$ *and assume that the level sets* $\Gamma(\lambda)$ *are* MV*-sets in* $\mathbb{C}$ *for all* $\lambda$ *in a neighbourhood of* $\mu_\alpha$. *Then we have for every* $\varepsilon > 0$ *small enough that*

$$d_F(Q_n(\alpha), Q(\alpha)) \leqslant F(\{x: |f(x) - \mu_\alpha| \leqslant \varepsilon\}) + M\varepsilon^{-1}(n^{-1/2}\tilde{q}_n(\alpha) - (F(Q_n(\alpha)) - F(Q(\alpha)))). \tag{3.6}$$

We call $n^{1/2}(F(Q_n(\alpha)) - F(Q(\alpha))$ the *generalized uniform quantile process*. The stochastic term in on the right-hand side of (3.6) equals $n^{-1/2}$ times the difference of the generalized and the generalized uniform quantile process. In the classical situation of one-dimensional quantiles this difference has been studied by Csörgö and Revesz (1978).

## 4. Bahadur–Kiefer approximations for $q_n$

As for the classical quantile process it is possible to derive Bahadur–Kiefer type approximations for the generalized quantile process. The given results in this sections actually generalize some of the classical results on Bahadur–Kiefer approximations.

**Theorem 4.1.** *Let $A \subset [0, 1]$ and $\lambda = Leb$. Suppose that $F$ is twice continuously differentiable such that* (3.1) *holds for $0 < \gamma \leqslant 1$ and that* (3.5) *is satisfied. In addition assume that all the level sets $\Gamma(\mu) = \{x: f(x) \geqslant \mu\}$, $\mu \geqslant 0$ lie in $\mathbb{C}$.*
(a) *If $\mathbb{C}$ is an n-deviation measurable $(v, m)$-constructible VC-class then for $\delta > \gamma/(2+\gamma)$*

$$\sup_{\alpha \in A} |\tilde{q}_n(\alpha) + v_n(Q(\alpha))| = O_{P^*}(n^{-\delta/2}) \quad \text{as } n \to \infty . \tag{4.1}$$

(b) *If $\mathbb{C}$ is such that* (3.3) *is satisfied for some $r > 0$ then*

$$\sup_{\alpha \in A} |\tilde{q}_n(\alpha) + v_n(Q(\alpha))| = O_{P^*}(n^{-\delta/(r)}) \quad \text{as } n \to \infty , \tag{4.2}$$

*where*

$$\delta_n(r) = \begin{cases} n^{-\delta} \text{ for } \delta > \gamma(1-r)/2(2+(1+r)\gamma) & \text{if } r < 1, \\ \log n & \text{if } r = 1, \\ n^{(r-1)/2(r+1)}, & \text{if } r > 1. \end{cases}$$

For a fixed regular level $\alpha$, where by definition $\gamma = 1$, we obtain from Theorem 4.1 that $|\tilde{q}_n(\alpha) + v_n(Q(\alpha))|$ is of the order $O_{P^*}(n^{-1/6+\varepsilon})$ for $n$-deviation measurable VC-class and $O_{P^*}(n^{-1/14+\varepsilon})$ for the class $\mathsf{C}^2(K)$. The latter follows since for $\mathsf{C}^d(K)$, $d \geqslant 2$, one has $r = r(d) = (d-1)/2$ (see above). Under mild conditions on the tail behaviour of $f$ the latter, and hence also the rate of approximations, can be extended to the class $\mathsf{C}^d$. (The result about the metric entropy of $\mathsf{C}^d$ can be found in Polonik, 1992). Note that Theorem 4.1 does not generalize the classical rates for Bahadur–Kiefer approximations for the one-dimensional quantile process, although the class $\{(-\infty, x] \in \mathbb{R}\}$ is a VC-class, but Theorem 4.2 does. See the end of Section 4 for an explanation.

The above theorem follows from Theorem 4.2 and (4.5) together with Theorem 3.1. Theorem 4.2 gives Bahadur–Kiefer approximations for the generalized empirical process for general $\lambda$. The given rates of approximation depend on the behaviour of the modulus of continuity of $\mathbb{C}$-indexed empirical process which is defined as

$$\omega_{v_n}(\delta) := \sup\{|v_n(C) - v_n(D)|;\ C, D \in \mathbb{C}, d_F(C, D) < \delta\}.$$

To formulate Theorem 4.2 we need the following assumption whose validity is discussed below:

(A8) *$V$ is differentiable in* $(0, 1)$ *with derivative* $V' = v > 0$.

**Theorem 4.2.** *Suppose that* (A0)–(A5) *and* (A8) *hold and let* $\eta \geqslant 0$.

*Part* I: *Suppose that for some* $\varepsilon > 0$ *small enough*

(i) $v$ is Lipschitz continuous in $[\eta - \varepsilon, 1 - \eta + \varepsilon] \cap [0, 1]$

(ii) $\inf_{\eta < \alpha < 1 - \eta} v(\alpha) > 0$

(iii) *for any choice of* $Q(\alpha)$ *the class* $\{Q(\alpha), \alpha \in [0, 1]\}$ *is an n-deviation measurable VC-class.*

(iv) $\|F_n - F\|_{\mathbb{C}} = O_{P^*}(h(n))$ *for a function with* $n^{-1/2} = O(h(n))$ *as* $n \to \infty$, *then*

$$\sup_{\eta < \alpha < 1 - \eta} |q_n(\alpha) + v_n(Q(\alpha))| = O_{P^*}(n^{1/2} h(n)) \text{ as } n \to \infty. \tag{4.3a}$$

*In addition let* $\{\delta_n\}$ *be a sequence of positive real numbers and assume that*

(v) $\alpha \to Q(\alpha)$, $\alpha \in [\eta - \varepsilon, 1 - \eta + \varepsilon] \cap [0, 1]$ *is Lipschitz continuous for* $d_F$.

(vi) $\sup_{\eta < \alpha < 1 - \eta} d_F(Q_n(\alpha), Q(\alpha)) = O_{P^*}(\delta_n)$ *as* $n \to \infty$.

(vii) $\omega_{v_n}(\delta_n) = O_{P^*}(g(\delta_n))$ *for a function* $g$ *with* $g(\delta_n) = g(\delta_n, n)$ *such that* $g(\delta_n) \to 0$ *and* $g(c\delta_n) = O(g(\delta_n))$ *as* $n \to \infty$ *for any* $c > 0$.

*If* $h(n) = O(\delta_n)$ *and* $n^{1/2}\, g(\delta_n) = O(\log n)$ *then we have*

$$\sup_{\eta < \alpha < 1 - \eta} |q_n(\alpha) + v_n(Q(\alpha))| = O_{P^*}(g(\delta_n)) \text{ as } n \to \infty. \tag{4.3b}$$

*Part* II: *Suppose that instead of* (i) *and* (ii) *of Part* I *we have that $v$ is differentiable in* $(0, 1)$ *with derivative $v'$ satisfying*

$$\sup_{0 < \alpha < 1} \alpha(1 - \alpha)|v'(\alpha)|(v(\alpha))^{-1} < c < \infty. \tag{4.4}$$

*Moreover, we assume that $v$ is monotone increasing in an interval* $(1 - \varepsilon, 1)$, $\varepsilon > 0$, *and either* $0 < \lim_{x \to 0} v(x) < \infty$ *or $v$ is monotone decreasing in an interval* $(0, \varepsilon)$, $\varepsilon > 0$. *If in addition the other assumptions of Part* I *are satisfied then* (4.3a) *and* (4.3b), *respectively, hold with* $\eta = 0$.

Functions $g$ satisfying condition (vii) are well known for certain classes $\mathbb{C}$. For VC-classes which satisfy some measurability condition we have $g(\delta_n) = (\delta_n \log n)^{1/2}$ (e.g. Pollard, 1984). If $\mathbb{C}$ satisfies (3.3), with $r > 0$, then it is known (Alexander, 1984) that if $n^{-2/(r+2)} \log n = O(\delta_n)$ then one can choose

$$g(\delta_n) = \begin{cases} \delta_n^{(1-r)/2} & \text{if } r < 1, \\ \log n & \text{if } r = 1, \\ n^{(r-1)/2(r+1)} & \text{if } r > 1. \end{cases} \tag{4.5}$$

**Remarks.** (i) Condition (4.4) has been used by Csörgö and Revesz (1978) to extend Kiefer's (1970) result for the one-dimensional quantile process (see also Corollary 4.3) to distributions with not necessarily compact support. Also Einmahl and Mason (1992) used this condition.

(ii) As in Theorem 3.1 we formulate the case $\eta = 0$ separately in Part II. The reason is given in the remarks after Theorem 3.1.

*Discussion of assumptions.* Assumption (iv) in Theorem 4.2 is trivially satisfied if the sets $Q(\alpha)$ are nested, because then $d_F(Q(\alpha), Q(\beta)) = |F(Q(\alpha)) - F(Q(\beta)) = |\alpha - \beta|$. This for example holds in the MV-case, provided the level sets are MV-sets. This also holds in the classical case of real-valued quantiles, i.e. for $\mathbb{C} = \{(-\infty, t], t \in \mathbb{R}\}$ and $\lambda((-\infty, t]) = t$. Examples where (A8) is satisfied are the following:

(a) In the case of the classical one-dimensional quantile process, i.e. for $\mathbb{C} = \{(-\infty, t]), t \in \mathbb{R}\}$ and $\lambda((-\infty, t]) = t$, we have for all $\alpha$ with $f(F^{-1}(\alpha)) > 0$ that $v(\alpha) = 1/f(F^{-1}(\alpha))$. Hence, if $f$ is differentiable, $v'(\alpha) = f'(F^{-1}(\alpha))/f^3(F^{-1}(\alpha))$.

(b) Consider the special situation of Section 3, i.e. let $\lambda = \text{Leb}$ and suppose that $\Gamma(\lambda) = \{x: f(x) \geqslant \lambda\} \in \mathbb{C}$ for all $\lambda > 0$. If $f$ has no flat parts, i.e. $F\{f = \mu\} = 0$ for all $\mu > 0$ (this is equivalent to (3.1) above), there exists for every $\alpha \in [0, 1]$ a unique $\mu_\alpha \geqslant 0$ such that $F(Q(\alpha) \Delta \Gamma(\mu_\alpha)) = 0$. $\mu_\alpha$ is defined through the equation $F(\Gamma(\mu_\alpha)) = \alpha$ (cf. Fig. 1) and we have $v(\alpha) = 1/\mu_\alpha$ for all $\alpha$. If in addition $f$ is continuously differentiable then $v'(\alpha) = (\mu_\alpha^3 \int_{\partial\Gamma(\mu_\alpha)} \|\text{grad} f(s)\|^{-1}\, ds)^{-1}$ where "ds" denotes the $(d-1)$-Hausdorff measure and $\partial\Gamma(\mu_\alpha) = \{x: f(x) = \mu_\alpha\}$. Note the analogy to (a), since $\mu_\alpha = f(\partial C(\alpha))$.

In the special case of classical real-valued quantiles we have $Q(\alpha) = (-\infty, F^{-1}(\alpha)]$ and $Q_n(\alpha) = (-\infty, F_n^{-1}(\alpha)]$. Hence, $d_F(Q_n(\alpha), Q(\alpha)) = |\int_{F_n^{-1}(\alpha)}^{F^{-1}(\alpha)} f(x)\, dx| \approx f(F^{-1}(\alpha))| F_n^{-1}(\alpha) - F^{-1}(\alpha)|$. The latter is the absolute value of the standardized one-dimensional quantile process (up to the factor $n^{1/2}$). Hence, a first application of Theorem 4.2, Part I, with $\delta_n = 1$ gives $d_F(Q_n(\alpha), Q(\alpha)) = O_{P^*}(n^{-1/2})$. Since the class $\mathbb{C} = \{(-\infty, t], t \in \mathbb{R}\}$ is a VC-class we have $\omega_{\nu_n}(\delta_n) = O_{P^*}(\delta_n^{-1/2}(\log \delta_n)^{1/2})$ and a second application of Theorem 4.2 gives an upper bound of $O_{P^*}(n^{-3/4}(\log n)^{1/2})$ which is the exact rate of approximation obtained by Kiefer (1970).

## 5. Testing for multimodality

The idea to construct tests for multimodality by means of a comparison of Lebesgue measures of MV-sets for different classes of sets $\mathbb{C}$ and $\mathbb{D}$ has been given in the introduction. Since in this section we simultaneously consider two different classes of sets $\mathbb{C}$ and $\mathbb{D}$, we add to our notation an index $\mathbb{C}$ or $\mathbb{D}$, respectively. For example, if the infimum in the definition of $V$ is extended over $\mathbb{D}$ we write $V_{\mathbb{D}}$ $v_{\mathbb{D}}$, $Q_{\mathbb{D}}(\alpha)$ instead of $V$, $v$ and $Q(\alpha)$, respectively. An analogous notation is used for the corresponding empirical versions. We also write $\nu_{n,\mathbb{D}}$ and $\nu_{n,\mathbb{C}}$ for the $\mathbb{C}$ and $\mathbb{D}$-indexed empirical process, respectively.

In general we consider the following test problem. Suppose that for two classes $\mathbb{C}$, $\mathbb{D}$ of measurable subsets of $\mathbb{R}^d$ with $\mathbb{C} \subset \mathbb{D}$ the minimum volume sets $Q_{\mathbb{C}}(\alpha)$ and $Q_{\mathbb{D}}(\alpha)$, in $\mathbb{C}$ and $\mathbb{D}$, respectively, are defined uniquely. Given $A \subset [0, 1]$ we consider the hypotheses

$$H_0\colon\ Q_{\mathbb{D}}(\alpha) \in \mathbb{C} \quad \text{for all } \alpha \in A$$

versus

$$H_1\colon\ Q_{\mathbb{D}}(\alpha) \in \mathbb{D}\backslash\mathbb{C} \quad \text{for some } \alpha \in A.$$

Let $V^S_{n,\mathbb{C}}(\alpha) = V_{n,\mathbb{C}}(\alpha)/V_{n,\mathbb{C}}(0.75)$ and $V^S_{n,\mathbb{D}}(\alpha) = V_{n,\mathbb{D}}(\alpha)/V_{n,\mathbb{D}}(0.75)$ denote scaled versions of $V_{n,\mathbb{C}}(\alpha)$ and $V_{n,\mathbb{D}}(\alpha)$, respectively. (Note that if $\mathbb{C}$ is the class of intervals, then $V_{n,\mathbb{C}}(0.75)$ denotes the interquartile range). As a test statistic for the above test problem we propose

$$T_{n,A}(\mathbb{C},\mathbb{D}) := \sup_{\alpha\in A}(V^S_{n,\mathbb{C}}(\alpha) - V^S_{n,\mathbb{D}}(\alpha)).$$

Proposition 5.1 shows that a test for $H_0$ versus $H_1$ based on $\mathrm{T}_{n,A}(\mathbb{C},\mathbb{D})$ is consistent: Let

$$T_A(\mathbb{C},\mathbb{D}) := \sup_{\alpha\in A}(V^S_{\mathbb{C}}(\alpha) - V^S_{\mathbb{D}}(\alpha)),$$

where $V^S_{\mathbb{C}}(\alpha) = V_{\mathbb{C}}(\alpha)/V_{\mathbb{C}}(0.75)$ and $V^S_{\mathbb{D}}(\alpha)$ is defined analogously with $\mathbb{C}$ replaced by $\mathbb{D}$.

**Proposition 5.1** (Consistency). *Let $\mathbb{C}$, $\mathbb{D}$ be GC-classes for $F$ with $\mathbb{C} \subset \mathbb{D}$. Assume that $V_{\mathbb{C}}$ and $V_{\mathbb{D}}$ are uniformly continuous on $A \subset [0,1]$, then as $n \to \infty$*

$$|\,T_{n,A}(\mathbb{C},\mathbb{D}) - T_A(\mathbb{C},\mathbb{D})| = \mathrm{O}_{\mathrm{P}^*}(1).$$

*Hence, under* $\mathrm{H}_0$ *as* $n \to \infty$

$$T_{n,A}(\mathbb{C},\mathbb{D}) = \mathrm{O}_{\mathrm{P}^*}(1).$$

**Theorem 5.2.** (Rates of convergence). *Let $\lambda = Leb$ and let $A = [\eta, 1-\eta]$, $\eta > 0$. Suppose that $F$ is twice continuously differentiable such that* (3.1) *holds for $0 < \gamma \leqslant 1$ and that* (3.5) *is satisfied. The following rates hold under* $\mathrm{H}_0$:

(a) *If $\mathbb{C}$ is an "n-deviation measurable" $(v,m)$-constructible VC-class then for $\delta > \gamma/(2+\gamma)$*

$$T_{n,A}(\mathbb{C},\mathbb{D}) = \mathrm{O}_{\mathrm{P}^*}(n^{-(1/2+\delta/2)}) \quad \textit{as } n \to \infty.$$

(b) *If* $\mathbb{C}$ is such that (3.3) is satisfied for some $r > 0$ then

$$T_{n,A}(\mathbb{C},\mathbb{D}) = \mathrm{O}_{\mathrm{P}^*}(n^{-(1/2+\delta(r))}) \quad \textit{as } n \to \infty$$

*where*

$$\delta_n(r) = \begin{cases} n^{-\delta} \textit{ for } \delta > \gamma(1-r)/2(2+(1+r)\gamma) & \textit{if } r < 1, \\ \log n & \textit{if } r = 1, \\ n^{(r-1)/2(r+1)} & \textit{if } r > 1. \end{cases}$$

*Under $H_1$ the rate in* (a) *and in* (b) *for $r < 1$ are both $n^{-1/2}$ which in this case are the exact rates.*

**Corollary 5.3.** *Suppose that the assumption of Theorem 5.2 hold. Assume that $f$ is unimodal with mode $x_0$ such that $\|\mathrm{grad} f(x)\| \neq 0$ for $x \neq x_0$ and that the level sets $\Gamma(\lambda) = \{x : f(x) \geqslant \lambda\}$ all lie in $\mathbb{C}$.*

(a) *Let $\mathbb{C}$ be the class of all intervals for $d = 1$ and the class of all ellipsoids for $d \geqslant 2$. If in addition $\mathbb{D}$ is a VC-class containing $\mathbb{C}$, then we have for every $\varepsilon > 0$ that as $n \to \infty$*

$$T_{n,A}(\mathbb{C}, \mathbb{D}) = \mathrm{O}_{\mathrm{P}^*}(n^{-2/3+\varepsilon}).$$

(b) *If $\mathbb{C}$ is the class of convex sets in $\mathbb{R}^d$ and $\mathbb{D}$ is k-constructible out of $\mathbb{C}$ then we have for every $\varepsilon > 0$ that as $n \to \infty$*

$$T_{n,A}(\mathbb{C}, \mathbb{D}) = \begin{cases} \mathrm{O}_{\mathrm{P}^*}(n^{-4/7+\varepsilon}) & \text{if } d = 2, \\ \mathrm{O}_{\mathrm{P}^*}(n^{-1/2}\log n) & \text{if } d = 3, \\ \mathrm{O}_{\mathrm{P}^*}(n^{-(2d-1)/(2d+4)}) & \text{if } d > 3. \end{cases}$$

Corollary 5.3 follows immediately from Theorem 5.2, because the assumptions assure that Theorem 5.2 can be applied with $\gamma = 1$.

## 6. Extensions and generalizations

In the proofs of the above results we do not explicitly use the i.i.d assumption, but only through the behaviour of the empirical process. Therefore the MV-approach can be transfered to situations where other empirical processes appear with similar properties as the usual empirical process used above. Such processes for example are the set-indexed partial sum process (this process appears in the regression context, see below) and the empirical spectral process which is used in spectral analysis.

*The regression problem*: Suppose that we have a nonparametric regression model on a regular grid $Y_i = r(i/n) + \varepsilon_i$, where $i \in \{0, 1, \ldots, n\}^d$, $r: [0, 1]^d \to [0, \infty)$ is the regression function and $\varepsilon_i$ are i.i.d. errors. Let $\mathbb{C}$ be a class of subsets of $[0, 1]^d$ and for $C \in \mathbb{C}$ let $R(C) = \int_C r(x)\,\mathrm{d}x$ and define

$$V(\alpha) = \inf\{\lambda(C): R(C) \geqslant \alpha,\ C \in \mathbb{C}\}, \quad 0 < \alpha < R, \tag{6.1}$$

and the corresponding empirical version

$$V_n(\alpha) = \inf\{\lambda(C): R_n(C) \geqslant \alpha,\ C \in \mathbb{C}\}, \quad 0 < \alpha < R. \tag{6.2}$$

where

$$R_n(C) = n^{-d} \sum_{i: i/n \in C} Y_i.$$

The minimizing set in (6.1) is a level set of the regression function $i$ if the level sets of $r$ lie in $\mathbb{C}$. Note that there are practical problem where one is interested in estimating level sets of a regression function (Messer, 1993). Similar results as in the previous sections can be proved by using the process $e_n(C) = n^{-d/2}(\sum_{i: i/n \in C} Y_i - R(C))$ instead of $v_n$. Under smoothness assumptions on $r$ one has

$$e_n(C) = n^{-d/2} \sum_{i: i/n \in C} \varepsilon_i + \mathrm{o}(1).$$

Set-indexed partial-sum processes of the form $n^{-d/2}\sum_{i: i/n \in C} \varepsilon_i$ have been studied (e.g. Bass and Pyke, 1984; Goldie and Greenwood, 1986; Alexander and Pyke, 1986). These results can be used to obtain results of the same type as given in the previous sections.

*Spectral analysis*: In spectral analysis one has a regression like situation with approximately independent "observations" if one considers the periodogram ordinates as observations $Y_i$ and the spectral density as regression function. Proceeding as above the empirical spectral process appears instead of the process $e_n$. See Dahlhaus (1988) for results on weak convergence of the empirical spectral process.

*Multivariate trimming*: Here we briefly indicate how the MV-approach can be used to rederive the limiting distribution of a certain process appearing in multivariate trimming as considered by Nolan (1992). This limiting distribution has already been derived by Nolan with different methods.

Nolan studies a method of multivariate trimming connected to quantiles of projections on $(d-1)$-dimensional hyperplanes. The trimming idea is to consider the intersection of all halfplanes in $\mathbb{R}^d$ which contain at least $(1-\alpha)$ percent of the data. The resulting convex set $C_n$ is called $\alpha$-trimmed region. Let $C$ denote the corresponding theoretical $\alpha$-trimmed region, i.e. the intersection of all halfplanes which contain $F$-mass at least $1-\alpha$. Nolan considered the following radius function on the unit sphere $\mathbb{S}^{d-1}$:

$$r_n(u) = \inf\{r \geqslant 0\colon ru \notin C_n\}, \quad u \in \mathbb{S}^{d-1}.$$

If the origin lies in $C_n$ then $ur_n(u)$ is an element of $\partial C_n$, otherwise $r_n(u) = 0$. Replace $C_n$ by $C$ to get the definition of $r_\alpha(u)$. Assume that $F$ is such that $C$ is non-empty and that (without loss of generality) 0 is an inner point of $C$. Let $\langle \cdot,\cdot \rangle$ denote the usual inner product on $\mathbb{R}^d$. Nolan showed that $r_n(u)$ has the same limiting distribution as

$$q_n^*(u) = q_n(v_\alpha(u))/\langle u, v_\alpha(u)\rangle$$

where $q_n(u)$ denotes the (one-dimensional) empirical $(1-\alpha)$-quantile of the projections $\langle u, X_i\rangle$, $i = 1, \ldots, n$, and $v_\alpha(u)$ is the outwarded (with respect to $C$) normal to $H(u)$, the supporting hyperplane to $C$ at $ru$. Now we indicate how the limiting distribution of $q_n(u)$ (and by that the limiting distribution of $q_n^*(u)$) can be determined by using generalized quantile processes. Let $\mathscr{H}(u)$ denote the class of all halfplanes $\{\langle x, u\rangle \leqslant c\}$, $c \in \mathbb{R}$, and for $H \in \mathscr{H}(u)$ let $r(H) = \inf\{r \geqslant 0\colon ru \notin H\}$. Then

$$q_n(u) = W_n(\alpha, u) = \inf\{r(H)\colon F_n(H) \geqslant 1-\alpha, H \in \mathscr{H}(u)\}$$

and

$$q_\alpha(u) = W(\alpha, u) = \inf\{r(H)\colon F(H) \geqslant 1-\alpha, H \in \mathscr{H}(u)\}.$$

Under appropriate smoothness assumptions the derivative of $W(\alpha, u)$ with respect to $\alpha$ is $\mathrm{w}(\alpha, \mathrm{u}) = (p_u(q_\alpha(u))^{-1}$ where $p_u$ is the density of the distribution of $\langle u, X\rangle$ under $F$. Analogous arguments as in the previous sections show that $n^{1/2}$ $(w(\alpha, v_\alpha(u))^{-1}$ $(q_n(v_\alpha(u)) - q_\alpha(v_\alpha(u)))$ can be approximated by $-n^{1/2}(F_n - F)(H(u))$ uniformly in $u \in S^{d-1}$. More precisely, if (among others) $\inf_{u \in S^{d-1}} w(\alpha, v_\alpha(u)) > 0$, then

$$\sup_{u \in S^{d-1}} |n^{1/2}(q_n(v_\alpha(u)) - q_\alpha(v_\alpha(u))) + n^{1/2} w(\alpha, v_\alpha(u))(F_n - F)(H(u))| = \mathrm{O}_\mathrm{P}(1).$$

Note that $q_\alpha^*(v_\alpha(u)) = \langle r_\alpha(u)u, v_\alpha(u)\rangle$. Hence, by definition of $q_n^*(u)$ it follows that $n^{1/2}$ $(q_n^*(u) - r_\alpha(u)$ has the same limiting distribution as $n^{1/2}$ $w(\alpha, v_\alpha(u))$ $\langle u, v_\alpha(u)\rangle^{-1}$

$(F_n - F)(H(u))$. Under smoothness assumptions the process $n^{1/2}$ $(F_n - F)(H(u))$ converges in distribution to a mean zero Gaussian process with covariance $c(u, v) = F(H(u) \cap H(v)) - F(H(u))F(H(v)) = F(H(u)H(v)) - (1 - \alpha)^2$. If the remainder terms can be controlled uniformly in $u$ (which should be possible under appropriate smoothness assumptions) then it follows that $n^{1/2}(q_n^*(u) - r_\alpha(u)$ converges in distribution to mean Gaussian process with covariance $c(u, v)\ g(u)^{-1} g(v)^{-1}$, where $g(u)^{-1} = w(\alpha, v_\alpha(u))\ \langle u, v_\alpha(u)\rangle^{-1}$.

## 7. Proofs

**Proof of Proposition 2.1.** Let $\bar{F}_n(t) = \sup\{F_n(C)\colon\ C \in \mathbb{C},\ \lambda(C) \leqslant V(t)\}$ and let $\bar{F}_n^{-1}(\alpha) = \inf\{t \in (0, 1)\colon \bar{F}_n(t) \geqslant \alpha\}$ be the generalized inverse of $\bar{F}_n(t)$. Einmahl and Mason (1992) showed (see their Lemma 3.1) that on the set where $V_n(0) < \infty$ (which has inner probability 1 by (A1)

$$V_n(\alpha) = V(\bar{F}_n^{-1}(\alpha)) \quad \text{for all } 0 < \alpha < 1 \tag{7.1}$$

Hence, $|V_n(\alpha) - V(\alpha)| = |V(\bar{F}_n^{-1}(\alpha)) - V(\alpha)|$ on a set with inner probability 1 and the continuity assumption on $V$ together with the fact that $\sup_{0 \leqslant \alpha \leqslant 1} |\bar{F}_n^{-1}(\alpha) - \alpha|^* \to 0$ a.s. (see Corollary 3.2 of Einmahl and Mason). □

**Proof of Proposition 2.2.** Let $\{\alpha_n\}$ be a sequence in $[\eta, 1 - \eta]$. We show that $d_F(Q_n(\alpha_n), Q(\alpha_n)) \to 0$ and $n \to \infty$ on a set with inner probability 1. Let $\alpha$ denote a limit point of $\{\alpha_n\}$. Then

$$d_F(Q_n(\alpha_n), Q(\alpha_n)) \leqslant d_F(Q_n(\alpha_n), Q(\alpha)) + d_F(Q(\alpha_n), Q(\alpha)).$$

First we show that $d_F(Q(\alpha_n), Q(\alpha))$ converges to zero. Let $Q$ be a limit point of $\{Q(\alpha_n)\}$ for $d_G$. Then there exists a subsequence $\{Q(\alpha_n')\}$ of $\{Q(\alpha_n)\}$converging to $Q$ in $d_G$ and hence also in $d_F$. It follows from the continuity of $V(\alpha) = \lambda(Q(\alpha))$ and the lower semicontinuity of $\lambda$ (assumption (A5)) that

$$\lambda(Q(\alpha)) = \liminf_{n'} (\lambda(Q(\alpha_{n'})) \geqslant \lambda(Q).$$

Since $|F(Q) - F(Q(\alpha_{n'}))| \leqslant d_F(Q(\alpha_{n'}), Q) \to 0$ and $F(Q(\alpha_{n'})) = \alpha_{n'} \to \alpha$ we have $F(Q) = \alpha$. The assertion now follows from the uniqueness of $Q(\alpha)$.

Now we show that $d_F(Q_n(\alpha_n), Q(\alpha))$ converges to zero on a set with inner probability 1. Let (for a fixed $\omega \in \Omega$) $R$ be a limit point of $\{Q_n(\alpha_n)\}$. It follows as above together with the consistency of $V_n$ (Proposition 2.1) that on a set with inner probability 1.

$$\lambda(Q(\alpha)) = \liminf_n \lambda(Q(\alpha_n)) = \liminf_n \lambda(Q_n(\alpha_n)) \geqslant \lambda(R).$$

It remains to show that $F(R) = \alpha$ for all $\omega \in A$ with $P_*(A) = 1$. From this the assertion follows by similar arguments as used above in proving that $d_F(Q(\alpha_n), Q(\alpha))$ converges to zero. In order to prove that $F(R) = \alpha$ for all $\omega \in A$ first note that $V$ is *strictly* monotone. This follows from (ii) and (A3). Together with the uniform convergence of $V_n$ to $V$ it follows that as $n \to \infty$

$$\sup_{\eta \leqslant \alpha \leqslant 1 - \eta} |F_n(Q_n(\alpha)) - \alpha|^* \to 0 \quad \text{a.s.} \tag{*}$$

Since also $\sup_{\eta \leqslant \alpha \leqslant 1-\eta} |F_n(Q_n(\alpha)) - F(Q_n(\alpha))| \leqslant \sup_{C \in \mathbb{C}} |F_n(C) - F(C)| \to 0$ a.s. the assertion follows. □

Before we give the proofs of Section 3 we first prove Theorem 4.2. The reason is that Theorem 4.2 will be used to prove the results of Section 3. The main technical result for deriving Theorem 4.2 is given by the following:

**Lemma 7.1.** *Suppose that* (A0), (A3) *and* (A8) *hold. Let* $A_0 \subset A \subset (0,1)$ *and* $c, d > 0$. *Suppose that the map* $\alpha \to Q(\alpha)$, $\alpha \in A$ *is Lipschitz continuous for* $d_F$ *with Lipschitz constant* $k \geqslant 1$. *For* $0 \leqslant \alpha \leqslant 1$ *and* $\delta > 0$ *let* $\alpha_n^{\pm} = \alpha - (F_n - F)(C(\alpha)) \pm n^{-1/2} \omega_{\nu_n}(k(c+d)\delta)$. *Then the following inequalities hold on the set* $B_n = \{\sup_{\alpha \in A_0} d_F(Q_n(\alpha), Q(\alpha)) < c\delta\} \cap \{\alpha_n^{\pm} \in A \,\forall \alpha \in A_0\} \cap \{\sup_{\alpha \in A_0} |\alpha_n^{\pm} - \alpha| < d\delta\}$ *for each* $\alpha \in A_0$:

$$q_n(\alpha) + \nu_n(Q(\alpha)) \leqslant v(\xi_n^+)v(\alpha)^{-1}\omega_{\nu_n}(k(c+d)\delta) - (v(\xi_n^+)v(\alpha)^{-1} - 1)\nu_n(Q(\alpha)), \tag{7.2a}$$

$$q_n(\alpha) + \nu_n(Q(\alpha)) \geqslant -v(\xi_n^-)v(\alpha)^{-1}\omega_{\nu_n}(k(c+d)\delta) - (v(\xi_n^-)v(\alpha)^{-1} - 1)\nu_n(Q(\alpha)), \tag{7.2b}$$

*where* $\xi_n^{\pm}$ *lie between* $\alpha_n^{\pm}(\delta)$ *and* $\alpha$, *respectively.*

**Proof of Lemma 7.1.** As in the proof of the consistency of $V_n(\alpha)$ (Proposition 2.1) the idea of the proof is to approximate $V_n(\alpha)$ through $V(\alpha_n)$ where $\alpha_n$ is random, so that we get rid of $V_n$. For each $\alpha$ we have

$$\begin{aligned} V_n(\alpha) &= \inf\{\lambda(C)\colon C \in \mathbb{C}, F_n(C) \geqslant \alpha\} \\ &= \inf\{\lambda(C)\colon C \in \mathbb{C}, F(C) \geqslant \alpha - (F_n - F)(C)\} \\ &= \inf\{\lambda(C)\colon C \in \mathbb{C}, F(C) \geqslant \alpha - (F_n - F)(Q(\alpha)) \\ &\qquad + ((F_n - F)(Q(\alpha)) - (F_n - F)(C))\}. \end{aligned}$$

Since the *inf* is attained at $Q_n(\alpha)$ and since $k \geqslant 1$ we have that on $B_n$ for each $\alpha \in A_0$

$$\begin{aligned} V_n(\alpha) = \inf\{&\lambda(C)\colon C \in \mathbb{C},\ d_F(C, Q(\alpha)) < k(c+d)\delta, \\ &F(C) \geqslant \alpha - (F_n - F)(Q(\alpha)) + ((F_n - F)(Q(\alpha)) - (F_n - F)(C))\}. \end{aligned}$$

Hence, it follows

$$\begin{aligned} &\inf\{\lambda(C)\colon C \in \mathbb{C}, F(C) \geqslant \alpha_n^-(\delta), d_F(C, Q(\alpha)) < k(c+d)\delta\} \\ &\quad \leqslant V_n(\alpha) \\ &\quad \leqslant \inf\{\lambda(C)\colon C \in \mathbb{C}, F(C) \geqslant \alpha_n^+, d_F(C, Q(\alpha)) < k(c+d)\delta\}. \end{aligned}$$

And since $d_F(Q(\alpha_n^{\pm}), Q(\alpha) \leqslant k|\alpha_n^{\pm} - \alpha| \leqslant kd\delta$ we have $V(\alpha_n^-(\delta)) \leqslant V_n(\alpha) \leqslant V(\alpha_n^+(\delta))$. Hence, the following inequality holds on $B_n$:

$$(v(\alpha))^{-1} n^{1/2}(V(\alpha_n^-(\delta)) - V(\alpha)) \leqslant q_n(\alpha) \leqslant (v(\alpha))^{-1} n^{1/2}(V(\alpha_n^+(\delta)) - V(\alpha)).$$

Inequalities (7.2a) and (7.2b) now follow easily by applying an one-term Taylor expansion to the right-hand side and the left-hand side of the last inequality, respectively. □

**Proof of Theorem 4.2.**

*Part* I: First we prove (4.3a). Since by assumption the class $\{Q(\alpha), \alpha \in [0, 1]\}$ is a VC-class it follows from standard results of empirical process theory (e.g. Shorak and Wellner 1986) that $\sup_{0<\alpha<1}|v_n(Q(\alpha))| = O_{P^*}(1)$. It remains to show that $\sup_{\eta<\alpha<1-\eta}|q_n(\alpha)| = O_{P^*}(n^{1/2}h(n))$. In order to do that we use the representation (7.1). It follows that on a set with inner probability 1 for all $0 < \alpha < 1$

$$q_n(\alpha) = v(\theta_n)v(\alpha)^{-1}n^{1/2}(\bar{F}_n^{-1}(\alpha) - \alpha) \tag{7.3}$$

for some $\theta_n$ lying between $\bar{F}_n^{-1}(\alpha)$ and $\alpha$. Since $\sup_{0\leqslant\alpha\leqslant 1}|\bar{F}_n^{-1}(\alpha) - \alpha|^* \to 0$ a.s. (cf. proof of Proposition 2.1) it follows from assumptions (i) and (ii) that $\sup_{\eta<\alpha<1-\eta}v(\theta_n)v(\alpha)^{-1} = O_{P^*}(1)$ as $n \to \infty$. Now we consider $n^{1/2}|\bar{F}_n^{-1}(\alpha) - \alpha|$. First note that $\sup_{0\leqslant\alpha\leqslant 1}n^{1/2}|\bar{F}_n^{-1}(\alpha) - \alpha| = \sup_{0\leqslant\alpha\leqslant 1}n^{1/2}|\bar{F}_n(\alpha) - \alpha|$ . For $\bar{F}_n(\alpha) - \alpha$ we have the following: $\bar{F}_n(\alpha)-\alpha \leqslant \sup\{(F_n - F)(C)\colon C \in \mathbb{C},\ \lambda(C) \leqslant V(\alpha)\}$, and hence, $\sup_{0<\alpha<1}n^{1/2}(\bar{F}_n(\alpha) - \alpha) \leqslant n^{1/2}\|F_n - F\|_{\mathbb{C}} = O_{P^*}(n^{1/2}h(n))$. Furthermore, $\bar{F}_n(\alpha) - \alpha \geqslant (F_n - F)(Q(\alpha))$, such that $-\sup_{0\leqslant\alpha\leqslant 1}n^{1/2}(\bar{F}_n(\alpha) - \alpha) = O_{P^*}(1)$. This proves (4.3a).

Now we show (4.3b) by means of Lemma 7.1 (Here the notation of Lemma 7.1 is used): Set $A = [\eta - \varepsilon, 1 - \eta + \varepsilon]\cap[0, 1]$ and $A_0 = [\eta, 1 - \eta]$. By definition of $\alpha_n^{\pm}$ we have $|\alpha_n^{\pm} - \alpha| \leqslant 3\|F_n - F\|_{\mathbb{C}}$. Hence, because of (A1) the (inner) probability of the set $\{\alpha_n^{\pm} \in A$ for all $\alpha \in A_0\}$ tends to one as $n$ tends to infinity. The same holds for $\xi_n^{\pm}$ instead of $\alpha_n^{\pm}$. It also follows that $\sup_{\alpha\in A_0}|\alpha_n^{\pm} - \alpha| = O_{P^*}(h(n))$. Because of the Lipschitz continuity of $\alpha \to Q(\alpha)$ (assumption (v)) the quantity $\sup_{\alpha\in A_0}d_F(Q\alpha_n^{\pm}), Q(\alpha))$ also is of order $O_{P^*}(h(n))$ and hence is $O_{P^*}(\delta_n)$.

Collecting all this it follows that for a given $\varepsilon > 0$ there exist constants $c, d > 0$ such that $P^*(B_n^c) < \varepsilon$ for $n$ large enough. Furthermore, on $B_n$ the factor $v(\xi_n^{\pm})\, v(\alpha)^{-1}$ is bounded uniformly in $\alpha \in A_0$. Hence, the first term on the right-hand side of (7.2a) gives the asserted order. It remains to consider the second term on the right-hand side of (7.2a). Let $K$ be the Lipschitz constant of $v$ on $A$, then we have on $B_n$

$$\begin{aligned}|(v(\xi_n^{+})v(\alpha)^{-1} - 1)v_n(Q(\alpha))| &\leqslant |Kv(\alpha)^{-1}(\xi_n^{+} - \alpha)v_n(Q(\alpha))|\\ &\leqslant K^*|\alpha_n^{+} - \alpha||v_n(Q(\alpha))|\end{aligned}$$

with $K^* = K\sup_{\alpha\in A_0}v(\alpha)^{-1}$. Note that $K^*$ is finite by assumption (ii). The last expression in the above sequence of inequalities of the order $O_{P^*}(h(n))$.

By using inequality (7.2b) instead of (7.2a) the argumentation for $-(q_n(\alpha) + v_n(Q(\alpha)))$ is completely analogous.

**Proof of Part II.** This proof is in principle the same as the proof of (4.3a) under the conditions of Part I. The only difference is how to show that $\sup_{0<\alpha<1}v(\theta_n)v(\alpha)^{-1}$ is stochastically bounded. Here the boundedness (in outer probability) follows from assumption (4.4). This has already been shown by Einmahl and Mason (1992).

(Actually they proved this only for $\alpha$ close to 1, but the case $\alpha$ close to 0 can be proved analogously since (4.4) also holds for $\alpha$ close to 0.)

Now we prove (4.3b). For short we write $\Delta_n(\alpha) = q_n(\alpha) + v_n(Q(\alpha))$. Again we use Lemma 7.1 (and the notation from this lemma) and we only give the upper bound for $\sup_{0<\alpha<1} \Delta_n(\alpha)$. By using (7.2b) instead of (7.2a) the upper bound for $-\sup_{0<\alpha<1} \Delta_n(\alpha)$ is completely analogous.

Fix $\varepsilon > 0$. Choose $c_\varepsilon > 0$ such that $P^*(\|F_n - F\|_{\mathcal{C}} > c_\varepsilon h(n)) < \varepsilon$ for $n$ large enough and let $k > 0$ denote the Lipschitz constant of $\alpha \to Q(\alpha)$. Let $\eta_n = 4kc_\varepsilon h(n)$ and define $A_n = [\eta_n, 1 - \eta_n]$. We split the interval (0, 1) into $(0, \eta_n)$, $A_n$ and $(1 - \eta_n, 1)$ and show that $\sup_{\alpha<\eta_n} \Delta(\alpha)$, $\sup_{\alpha>1-\eta_n} \Delta_n(\alpha)$ and $\sup_{\alpha\in A_n} \Delta_n(\alpha)$ are of the asserted order.

First we consider the case $\alpha \in A_n$. Lemma 7.1 will be applied with $A_0 = A_n$ and $A = (0, 1)$. Let $d_\varepsilon > 0$ be such that $P^*(\sup_{\alpha\in A} d_F(Q_n(\alpha), Q(\alpha)) > d_\varepsilon \delta_n) < \varepsilon$ for $n$ large enough. Since $h(n) = O(\delta_n)$ there exists a constant $c_1 > 0$ with $h(n) \leqslant c_1 \delta_n$. We show that for the set $B_n$ as defined in Lemma 7.1 with $c = d_\varepsilon$ and $d = kc_1 c_\varepsilon \delta_n$ we have

$$P^*(B_n^c) < \varepsilon \quad \text{as } n \to \infty. \tag{7.6}$$

Note that $\sup_{\alpha\in A_n} d_F(Q\alpha_n^+), Q(\alpha)) \leqslant k \sup_{\alpha\in A_n} |\alpha_n^+ - \alpha| \leqslant 3k\|F_n - F\|_{\mathcal{C}}$. Hence, $P^*(\sup_{\alpha\in A_n} d_F(Q(\alpha_n^+), Q(\alpha)) > kc_1 c_\varepsilon \delta_n) < \varepsilon$ for $n$ large enough. Furthermore, we have $P_*(\alpha_n^+ \in (0, 1) \forall \alpha \in A_n) \geqslant 1 - \varepsilon$ for $n$ large enough. This follows from $P^* \sup_{\alpha\in A_n} |\alpha_n^+ - \alpha| > 3c_\varepsilon h(n)) < \varepsilon$ and our choise of $\eta_n$. Hence (7.6) follows.

On $B_n$ we know that (7.2a) holds. Therefore the assertion follows if we have shown that

$$\sup_{\alpha\in A_n} v(\xi_n^+) v(\alpha)^{-1} = O_{P^*}(1), \tag{7.7}$$

$$\sup_{\alpha\in A_n} |(v(\xi_n^+) v(\alpha)^{-1} - 1) v_n(Q(\alpha))| = O_{P^*}(g(\delta_n)). \tag{7.8}$$

We first prove (7.8). In order to do this we first rewrite the term $(v(\xi_n^+) v(\alpha)^{-1} - 1)$. Since $v$ is assumed to be differentiable we have on $B_n$

$$\begin{aligned} v(\xi_n^+) v(\alpha)^{-1} - 1 &= v'(\theta_n^+) v(\alpha)^{-1} (\xi_n^+ - \alpha) \\ &= [\theta_n^+(1 - \theta_n^+) v'(\theta_n^+) v(\theta_n^+)^{-1}][v(\theta_n^+) v(\alpha)^{-1}] \\ &\quad [\theta_n^+(1 - \theta_n^+)]^{-1} (\xi_n^+ - \alpha), \end{aligned}$$

where $\theta_n^+$ lies between $\xi_n^+$ and $\alpha$. By assumption (4.4) the first term in the last line is bounded on $B_n$ uniformly $\alpha \in A_n$. Hence, to prove (7.8) it remains to show

$$\sup_{\alpha\in A_n} (v(\theta_n^+) v(\alpha)^{-1})(\theta_n^+(1 - \theta_n^+))^{-1} \alpha(1 - \alpha) = O_{P^*}(1), \tag{7.9}$$

$$\sup_{\alpha\in A_n} |(\alpha(1 - \alpha))^{-1} (\xi_n^+ - \alpha) v_n(Q(\alpha))| = O_{P^*}(g(\delta_n)). \tag{7.10}$$

It turns out that the proof of (7.9) is similar to the proof of (7.7). Therefore the proof of (7.9) will be given below together with the proof of (7.7). Now we prove (7.10). Since $\xi_n^+$ lies between $\alpha_n^+$ and $\alpha$ we have

$$|\xi_n^+ - \alpha| \leqslant |\alpha_n^+ - \alpha| \leqslant |(F_n - F)(Q(\alpha))| + |n^{-1/2} \omega_{v_n}(\delta_n)|.$$

Hence, (7.10) follows from (7.11) and (7.12), where

$$\sup_{\alpha\in A_n} n^{1/2}(\alpha(1-\alpha))^{-1}|(F_n-F)(Q(\alpha))|^2 = O_{P^*}(g(\delta_n)), \tag{7.11}$$

$$\sup_{\alpha\in A_n}(\alpha(1-\alpha))^{-1}|(F_n-F)(Q(\alpha))| = O_{P^*}(1). \tag{7.12}$$

Note that by assumption $F(Q(\alpha)) = \alpha$ and that $\{Q(\alpha), \alpha\in[0,1]\}$ is a VC-class. Hence, (7.11) and (7.12) follow from that for VC-classes $\mathbb{C}$

$$\sup_{\eta_n\leqslant F(C)\leqslant 1-\eta_n} q_1(F(C))^{-1}|v_n(C)| = O_{P^*}(n^{1/4}g(\delta_n)^{1/2}, \tag{7.13}$$

$$\sup_{\eta_n\leqslant F(C)\leqslant 1-\eta_n} q_2(F(C))^{-1}|v_n(C)| = O_{P^*}(n^{1/2}), \tag{7.14}$$

where $q_1(t) = (t(1-t))^{1/2}$ and $q_2(t) = t(1-t)$. However, since by assumption $n^{1/2}g(\delta_n) = O(\log n)$ Eq. (7.13) follows from Alexander (1985, Corollary 1.6 (i)). Eq. (7.14) directly follows from Alexander (1985), Corollary 1.6(ii). To finish the proof of $\sup_{\alpha\in A_n}\Delta_n(\alpha) = O_{P^*}(g(\delta_n))$ it remains to show (7.7) and (7.9). First note that

$$v(s)v(t)^{-1} \leqslant ([(s\vee t)(1-(s\wedge t))][s\wedge t)(1-(s\vee t))]^{-1})^c. \tag{7.15}$$

This can be obtained from assumption (4.4) by easy analysis. (See for example Shorack and Wellner (1986, p. 644).) By means of (7.15) we get that (7.7) and (7.9) follow from

$$\sup_{\alpha<1-\eta_n}(1-\alpha)(1-\xi_n^+)^{-1} = O_{P^*}(1) \quad\text{and}\quad \sup_{\alpha>\eta_n}\alpha\xi_n^{+\,-1} = O_{P^*}(1) \tag{7.16}$$

and that the same holds with $\xi_n^+$ replaced by $\theta_n^+$. This is because (7.15) and (7.16) imply that $\sup_{\alpha\in A_n}(v(\xi_n^+)v(\alpha)^{-1}) = O_{P^*}(1)$ and that $\sup_{\alpha\in A_n}(v(\theta_n^+)v(\alpha)^{-1}) = O_{P^*}(1)$. (7.16) can be seen as follows. Since $\xi_n^+$ and $\theta_n^+$ lie between $\alpha_n^+$ and $\alpha$ we have

$$\alpha - |\alpha_n^+ - \alpha| \leqslant \xi_n^+ \leqslant \alpha + |\alpha_n^+ - \alpha|$$

and the same holds for $\theta_n^+$ instead of $\xi_n^+$. Hence, it suffices to show

$$P^*\left(\sup_{\alpha\in A_n}|\alpha_n^+ - \alpha|^*(\alpha(1-\alpha))^{-1} > 1/2\right) \to 0 \quad \text{as } n\to\infty. \tag{7.17}$$

Note that $\sup_{\alpha\in A_n} n^{-1/2}\omega_{v_n}(\delta_n)(\alpha(1-\alpha))^{-1} \leqslant 2^{-1}\eta_n^{-1}n^{-1/2}\omega_{v_n}(\delta_n) = O_{P^*}(g(\delta_n)) = O_{P^*}(1)$. This together with (7.14) gives (7.17).

To complete the proof of the theorem we have to show that $\sup_{\alpha\leqslant\eta_n}|\Delta_n(\alpha)|$ and $\sup_{\alpha\geqslant 1-\eta_n}|\Delta_n(\alpha)|$ also are of the order $O_{P^*}(g(\delta_n))$. We first discuss the idea for the case $\alpha \geqslant 1-\eta_n$. By definition of $g$ we have $\sup_{\alpha\geqslant 1-\eta_n}|v_n(Q(\alpha))| = O_{P^*}(g(\eta_n)) = O_{P^*}(g(\delta_n))$. Hence it remains to show that $\sup_{\alpha\geqslant 1-\eta_n}|q_n(\alpha)|$ is of the same order. We again make a splitting and first consider the region $\{1-\gamma_n \geqslant \alpha \geqslant 1-\eta_n\}$ for an appropriate sequence $\gamma_n \leqslant \eta_n$. Finally, we will show the assertion for $\{1 > \alpha \geqslant 1-\gamma_n\}$.

We choose $\gamma_n = 2C_\varepsilon n^{-1/2}g(\delta_n)$, where $C_\varepsilon > 0$ is such that $P^*(\omega_{v_n}(\delta_n) \geqslant C_\varepsilon g(\delta_n)) \leqslant \varepsilon$. Then the above arguments for the case $\{\alpha\in A_n\}$ can be carried through analogously for $\{1-\gamma_n \geqslant \alpha \geqslant 1-\eta_n\}$ and also for $\{\gamma_n \leqslant \alpha \leqslant \eta_n\}$. It remains to show $\sup_{\alpha\geqslant 1-\gamma_n}|q_n(\alpha)| = O_{P^*}(g(\delta_n))$ and $\sup_{\alpha\leqslant\gamma_n}|q_n(\alpha)| = O_{P^*}(g(\delta_n))$.

We again first concentrate on the case $\alpha$ close to 1. To show that $\sup_{\alpha \geqslant 1-\gamma_n} |q_n(\alpha)|$ is of the asserted order we again use the representation (7.1). This leads to

$$q_n(\alpha) = n^{1/2} \int_\alpha^{\bar{F}_n^{-1}(\alpha)} v(s) v(\alpha)^{-1} \, ds. \tag{7.18}$$

If $\alpha \geqslant \bar{F}_n^{-1}(\alpha)$ then the integrand is bounded by 1 and $q_n(\alpha) \leqslant n^{1/2}(\alpha - \bar{F}_n^{-1}(\alpha))$. It follows that in this case $\sup_{\alpha \geqslant 1-\gamma_n} |q_n(\alpha)| \leqslant n^{1/2} \sup_{\alpha \geqslant 1-\gamma_n} (\alpha - \bar{F}_n^{-1}(\alpha)) = n^{1/2} \sup_{\alpha \geqslant \bar{F}_n^{-1}(1-\gamma_n)} (\bar{F}_n(\alpha) - \alpha)$. To estimate the last term we use the following fact: for $t \in [0, 1]$ we have up to an exceptional set with outer probability tending to 0 that

$$\sup_{\alpha \geqslant t}(\bar{F}_n(\alpha) - \alpha) \leqslant \sup_{F(A) \geqslant t - h(n)\log n} (F_n(A) - F(A)), \tag{7.19}$$

where here and in the following the sup is extended over $A \in \mathbb{C}$. Using this and the fact that $\sup_{\alpha \in [0,1]} |\bar{F}_n^{-1}(\alpha) - \alpha| = O_{P^*}(n^{-1/2})$, which follows from Corollary 3.1 of Einmahl and Mason (1992), gives us

$$\begin{aligned}
\sup_{\alpha \geqslant 1-\gamma_n} |q_n(\alpha)| &\leqslant n^{1/2} \sup_{F(A) \geqslant 1-\gamma_n - O_P^*(n^{-1/2})} (F_n(A) - F(A)) \\
&\leqslant n^{1/2} \sup_{F(A) \geqslant 1-\gamma_n - O_P^*(n^{-1/2}) - h(n)\log n} (F_n(A) - F(A)) \\
&= O_{P^*}(g(h(n)\log n)) = O_{P^*}(g(\delta_n)).
\end{aligned}$$

It remains to show (7.19). For any sequence $\beta_n$ we have

$$\begin{aligned}
(\bar{F}_n(\alpha) - \alpha) &\leqslant (\sup\{F_n(A) \colon \alpha - \beta_n \leqslant F(A) \leqslant \alpha\} - \alpha) \\
&\quad \vee (\sup\{F_n(A) \colon \alpha - \beta_n > F(A)\} - \alpha). \\
&\leqslant \left( \sup_{F(A) \geqslant \alpha - \beta_n} (F_n(A) - F(A)) \right) \\
&\quad \vee \left( \sup_{A \in \mathbb{C}} (F_n(A) - F(A)) - \beta_n \right).
\end{aligned}$$

Since $\sup_{\alpha \in \mathbb{C}}(F_n(A) - F(A) = O_{P^*}(h(n))$ the assertion follows by choosing $\beta_n = h(n)\log n$. For the case $\alpha < \bar{F}_n^{-1}(\alpha)$ we again start with (7.18) and use (7.15). This gives for $\alpha \geqslant \frac{1}{2}$ and $c \neq 1$

$$\begin{aligned}
q_n(\alpha) &\leqslant 2^c n^{1/2}(1-\alpha)^c \int_\alpha^{\bar{F}_n^{-1}(\alpha)} (1-s)^{-c} \, ds \\
&\leqslant 2^c(1-c)^{-1} n^{1/2}(1-\alpha)^c[(1-\bar{F}_n^{-1}(\alpha))^{1-c} - (1-\alpha)^{1-c}] \\
&= 2^c(1-c)^{-1} n^{1/2}[(1-\alpha)((1-\alpha)/(1-\bar{F}_n^{-1}(\alpha)))^{c-1} - (1-\alpha)].
\end{aligned}$$

For $c < 1$ we have of course $((1-\alpha)/(1-\bar{F}_n^{-1}(\alpha)))^{c-1} \leqslant 1$ and the assertion follows by choice of $\gamma_n$. For $c > 1$ we use the fact that $\sup_{\alpha \in [0,1]}((1-\alpha)/(1-\bar{F}_n^{-1}(\alpha)))$ is bounded in probability (Lemma 3.3, Einmahl and Mason, 1992).

Analogous arguments can be used for $c = 1$ where we have $\int_\alpha^{\bar{F}_n^{-1}(\alpha)}(1-s)^{-c} \, ds = \log((1-\alpha)/(1-\bar{F}_n^{-1}(\alpha)))$.

The case $\alpha \leqslant \gamma_n$ follows the same line as the just considered case $\alpha \geqslant 1 - \gamma_n$. Therefore we only give a brief outline. If $v$ is bounded in a neighbourhood of 0 and bounded away from zero then we just need to show the analog of (7.19), and then the above arguments can be carried through. This analog is that for $t \in [0, 1]$ we have

$$\sup_{\alpha \leqslant t}(\bar{F}_n(\alpha) - \alpha) \leqslant \sup_{F(A) \leqslant t} (F_n(A) - F(A))$$

where on the right-hand side again the sup is extended over $A \in \mathbb{C}$. This fillows almost directly from the definition of $\bar{F}_n$.

If $v$ is unbounded and monotone decreasing near 0, then similar arguments as above show that the crucial situation here is $\bar{F}_n^{-1}(\alpha) \leqslant \alpha$ and $c \geqslant 1$. It follows from (7.18) and (7.15) that for $\alpha \leqslant 1/2$

$$q_n(\alpha) \leqslant 2^c(c-1)^{-1} n^{1/2}[\alpha(\alpha/\bar{F}_n^{-1}(\alpha))^{c-1} - \alpha]. \tag{7.20}$$

Analoguous arguments as used above for the case $\alpha \geqslant 1 - \gamma_n$ can now be used by just replacing $1 - \alpha$ and $1 - \bar{F}_n^{-1}(\alpha)$ through $\alpha$ and $\bar{F}_n^{-1}(\alpha)$, respectively. Note that Lemma 3.3 of Einmahl and Mason can under the stronger condition (4.4) easily be adapted to show

$$\sup_{0 \leqslant \alpha \leqslant 1} \alpha/\bar{F}_n^{-1}(\alpha) = O_{p^*}(1) \text{ as } n \to \infty. \tag{7.21}$$

**Proof of Lemma 3.2.** Note that $v(\alpha) = \mu_\alpha^{-1}$ for $\alpha < 1$, where $\mu_\alpha$ has been defined in Section 3. Hence, the stochastic term of the right-hand side of (3.6) (i.e. $n^{-1/2}$ times the difference of the generalized and the generalized uniform empirical process) can be rewritten as

$$n^{-1/2}\tilde{q}_n(\alpha) - (F(Q_n(\alpha)) - F(Q(\alpha))) = H_{\mu_\alpha}(Q(\alpha)) - H_{\mu_\alpha}(Q_n(\alpha)), \tag{7.22}$$

where $H_\mu(C) = (F - \lambda Leb)(C) = \int_C (f(x) - \mu)\,dx$, $\mu \geqslant 0$. The functional $H_{\mu_\alpha}$ is maximized by $\Gamma(\mu_\alpha)$. Since $\Gamma(\mu_\alpha) \in \mathbb{C}$ is a minimum volume set it has the same Lebesgue measure as $Q(\alpha)$. By definition $F(\Gamma(\mu_\alpha)) = F(Q(\alpha)) = \alpha$. Hence $H_{\mu_\alpha}$ is also maximized by $Q(\alpha)$. Note that this means that the difference of the generalized and the generalized uniform empirical process is nonnegative. (The maximal value $H_\mu(\Gamma(\mu))$ as a function in $\mu$ is called *excess mass* and has been considered in Hartigan, 1987; Müller and Sawitzki, 1989: Nolan, 1989; Polonik, 1992, 1995a, 1995b). Now, for any $\mu_\alpha > 0$ and any $\varepsilon > 0$ we have

$$\begin{aligned} H_{\mu_\alpha}(Q_n(\alpha)) - H_{\mu_\alpha}(Q_n(\alpha)) = H_{\mu_\alpha}(\Gamma(\mu_\alpha)) - H_{\mu_\alpha}(Q_n(\alpha)) &= \int_{D_n(\alpha)} |f(x) - \mu_\alpha|\,dx \\ &\geqslant \varepsilon Leb(D_n(\alpha) \cap \{x: |f(x) - \mu_\alpha| > \varepsilon), \end{aligned} \tag{7.23}$$

where for short $D_n(\alpha) = Q_n(\alpha) \Delta \Gamma(\mu_\alpha)$. *Hence*, (7.25) and (7.26) give for any $\varepsilon > 0$

$$\begin{aligned} F(D_n(\alpha)) &= F(D_n(\alpha) \cap \{x: |f(x) - \mu_\alpha| \leqslant \varepsilon\}) + F(D_n(\alpha) \cap \{x: |f(x) - \mu_\alpha| > \varepsilon\}) \\ &\leqslant F(\{x: |f(x) - \mu_\alpha| \leqslant \varepsilon\}) + M\ Leb(D_n(\alpha) \cap \{x: |f(x) - \mu_\alpha| > \varepsilon\}) \\ &\leqslant F(\{x: |f(x) - \mu_\alpha| \leqslant \varepsilon\}) + M\varepsilon^{-1}(H_{\mu_\alpha}(Q(\alpha)) - H_{\mu_\alpha}(Q_n(\alpha))). \quad \square \end{aligned}$$

**Proof of Theorem 3.1.** First note, that from (A7) it follows that all $\alpha \in (0, 1)$

$$n^{-1/2}\tilde{q}_n(\alpha) - (F(Q_n(\alpha)) - F(Q(\alpha))) = [n^{-1/2}\tilde{q}_n(\alpha) + (F_n - F)(Q_n(\alpha))] + R_n,$$

where $R_n^* = \mathrm{O}(1/n)$ (uniformly in $\alpha$). Hence,

$$\begin{aligned}\sup_{\eta<\alpha<1-\eta} d_F((Q_n(\alpha)\Delta Q(\alpha)) &\leqslant \mathrm{O}(\varepsilon_n^{\gamma}) + M\varepsilon_n^{-1}[n^{-1/2}\tilde{q}_n(\alpha) - (F(Q_n(\alpha)) - F(Q(\alpha)))] \\ &\leqslant \mathrm{O}(\varepsilon_n^{\gamma}) + M\varepsilon_n^{-1}[n^{-1/2}\tilde{q}_n(\alpha) + (F_n - F)(Q(\alpha)))] \\ &\quad - [(F_n - F)(Q(\alpha)) - (F_n - F)(Q_n(\alpha)))] + R_n]. \qquad (7.24)\end{aligned}$$

Below we use the following fact: On the set $A_\delta = \{\sup_{\eta<\alpha<1-\eta} d_F((Q_n(\alpha)\Delta Q(\alpha)) \leqslant \delta\}$, $\delta > 0$ we have for the stochastic term on the right-hand side of (7.24)

$$\begin{aligned}&\sup_{\eta<\alpha<1-\eta} [[n^{-1/2}\tilde{q}_n(\alpha) + (F_n - F)(Q(\alpha)))] \\ &\qquad - [(F_n - F)(Q(\alpha)) - (F_n - F)(Q_n(\alpha)))] + R_n] \\ &= \mathrm{O}_{\mathrm{P}^*}(\omega_{v_n}(\delta_n)).\end{aligned}$$

Since the sets $Q(\alpha)$ are nested the class $\{Q(\alpha), \alpha \in [0, 1]\}$ is a VC-class (of index 1). Furthermore, we have $v(\alpha) = 1/\mu_\alpha$ such that (i) and (ii) of Theorem 4.2 follow from (i) and (ii) of Theorem 3.1. Theorem 4.2 now shows, that the two stochastic terms in brackets on the right-hand side of (7.24) are of the same order. Therefore, an application of (4.3a) (with $h_n = n^{-1/2}$) gives for any sequence $\{\varepsilon_n\}$ of real numbers

$$\sup_{\eta<\alpha<1-\eta} d_F((Q_n(\alpha)\Delta Q(\alpha)) \leqslant \mathrm{O}(\varepsilon_n^{\gamma}) + M\varepsilon_n^{-1}\mathrm{O}_{\mathrm{P}^*}(n^{-1/2}) + R_n.$$

Choose $\varepsilon_n = \mathrm{O}(n^{-1/2(1+\gamma)})$ to balance the two terms on the right-hand side. The resulting rate of convergence is $\mathrm{O}_{\mathrm{P}^*}(n^{-\gamma^*/2})$, where for short $\gamma^* = \gamma/(1+\gamma)$. Now we split the proof. First we proof Part I(a).

An application of (4.3b) with $\delta_n = n^{-\gamma^*/2}$ gives a rate of convergence for both of the two stochastic terms in brackets in (7.24) of $\delta_n = (n^{-\gamma^*/2}\log n)^{1/2}$ since for VC-classes (vii) of Theorem 4.2 holds with $g(\delta_n) = (\delta_n \log n)^{1/2}$ (e.g. Pollard, 1984). Choose $\varepsilon_n = \mathrm{O}((n^{-1/2}(n^{-\gamma^*/2}\log n)^{1/2})^{1/(1+\gamma)})$ to balance the stochastic and the non stochastic term on the righthand side of (7.24). This leads to the new rate $\mathrm{O}_{\mathrm{p}^*}(n^{-\gamma^*/2(1+\gamma^*/2)}$ $(\log n)^{\gamma^*/2})$. Iterating these arguments $k$-times leads to the rate $n^{-r_k}(\log n)^{r_k-1}$ with $r_k = \gamma^*/2\sum_{j=0}^{k-1}(\gamma^*/2)^j$. We can do this iteration arbitrarily but finitely often. Since $r_\infty = (\gamma^*/2)/(1-\gamma^*/2) = \gamma/(2+\gamma)$ equality (3.2) follows.

The proof of (3.4) is quite similar. Here functions $g$ satisfying (vii) of Theorem 4.2 are given in (4.5). Using these, analogous arguments as for VC-classes gives the asserted rates. For $r < 1$ the rate after $k$ iterations is $\mathrm{O}_{\mathrm{P}^*}(n^{-r_k})$ with $r_k = \gamma^*/2$ $\sum_{j=0}^{k-1}(\gamma^*(1+r)/2)^j$. Since $r_\infty = \gamma/(2+(1+r)\gamma)$ the assertion follows. For $r \geqslant 1$ iteration is useless since $\delta_n$ does not enter the function $g$ and we are finished after the first step. □

**Proof of Proposition 5.1.** It follows from Proposition 2.1 that

$$\sup_{\alpha\in A}|(V^S_{n,\mathbb{C}}(\alpha) - V^S_{n,\mathbb{D}}(\alpha)) - (V^S_{\mathbb{C}}(\alpha) - V^S_{\mathbb{D}}(\alpha))| = \mathrm{O}_{\mathrm{P}^*}(1)$$

Since in addition under $H_0$ we have for all $\alpha \in A$ that $V^S_{\mathbb{C}}(\alpha) = V^S_{\mathbb{D}}(\alpha)$ the assertion follows. □

**Proof of Theorem 5.2.** Note that (3.5) is a version of assumption (4.4) (see Remark after Corollary 3.2). Hence, it follows from the proof of Theorem 4.2 that

$$n^{1/2}(V_{n,\mathbb{D}}(\alpha) - V_{\mathbb{D}}(\alpha)) = v_{\mathbb{D}}(\alpha)\, v_n(Q_{\mathbb{D}}(\alpha)) + v_{\mathbb{D}}(\alpha)\omega_{v_{n,\mathbb{D}}}(\delta_n) + R_{n,\mathbb{D}}, \tag{7.25}$$

where $\delta_n$ is the rate of convergence of $\sup_{\alpha\in A} d_F(Q_n(\alpha),Q(\alpha))$ and $R_{n,\mathbb{D}}$ is a remainder term which is of lower order (uniformly in $\alpha$). The same holds for $\mathbb{C}$ replaced for $\mathbb{D}$. Under $H_0$ the first terms on the right-hand side in (7.28) are the same for $\mathbb{C}$ and $\mathbb{D}$, respectively, so that

$$|n^{1/2}(V_{n,\mathbb{C}}(\alpha) - V_{\mathbb{C}}(\alpha)) - (V_{n,\mathbb{D}}(\alpha) - V_{\mathbb{D}}(\alpha))| \leqslant 2v_{\mathbb{D}}(\alpha)\omega_{v_{n,\mathbb{D}}}(\delta_n) + R_{n,\mathbb{D}} + R_{n,\mathbb{C}}.$$

Since $v_{\mathbb{D}}(\alpha) = 1/\mu_\alpha$ and $\alpha$ is bounded away from 1 (and $f$ is bounded) we have $\sup_{\alpha\in A} v_{\mathbb{D}}(\alpha) = \sup_{\alpha\in A} 1/\mu_\alpha < \infty$. The rate $\delta_n$ we get from Theorem 3.1. This together with the behaviour of $\omega_{v_{n,\mathbb{D}}}(\delta_n)$ which has been derived by Alexander (1984) (Correction: Alexander, 1985) (cf. Section 4 after Theorem 4.2) gives the assertion for a test statistics analogous to $T_{n,A}(\mathbb{C},\mathbb{D})$ constructed out of $V_{n,\mathbb{D}}$ and $V_{n,\mathbb{C}}$ instead of their standardized versions $V^S_{n,\mathbb{D}}$ and $V^S_{n,\mathbb{C}}$, respectively. It remains to show that the difference of these test statistics is at least of the same order, but this follows by repeatedly using elementary reformulations of the type $V^S_{n,\mathbb{D}}(\alpha) - V^S_{\mathbb{D}}(\alpha) = (V_{\mathbb{D}}(0.75)/V_{n,\mathbb{D}}(0.75) - 1)\; V_{\mathbb{D}}(0.75)^{-1}(V_{n,\mathbb{D}}(\alpha) + V_{\mathbb{D}}(0.75)^{-1}(V_{n\mathbb{D}}(\alpha) - V_{\mathbb{D}}(\alpha))$.

Under $H_1$ the first order terms in (7.28) for $\mathbb{C}$ and $\mathbb{D}$, respectively, do not cancel. Hence the asserted rate of $\mathrm{O}_{\mathrm{P}^*}(n^{-1/2})$ follows. □

## References


Alexander, K.S., 1984. Probability inequalities for empirical processes and the law of the iterated logarithm. Ann. Probab. 12, 1041–1067; Correction: (1987) Ann. Probab. 15, 428–430.

Alexander, K.S., 1985. Rates of growth for weighted empirical processes. In: Le Cam, L.M., Olshen, R.A. (eds.), Proc. Berkeley Conf. Honor of Jerzy Neyman and Jack Kiefer, vol. 11. Wadsworth, Belmont, CA.

Alexander, K.S., Pyke, R., 1986. A uniform central limit theorem for set-indexed partial-sum processes with finite variance. Ann. Probab. 14, 582–597.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rodgers, W.H., Tukey, J.W., 1972. Robust Estimation of Location: Survey and Advances. Princeton Univ. Press, Princeton, NJ.

Bahadur, R.R. 1966. A note on quantiles in large samples. Ann. Math. Statist. 37, 577–580.

Bass, R.F., Pyke, R., 1984. Functional law of the iterated logarithm and uniform central limit theorem for partial-sum processes indexed by sets. Ann. Probab. 12, 13–34.

Csörgö, M., Révész, P., 1978. Strong approximations of the quantile process. Ann. Statist. 6, 882–894

Cuevas, A., 1990. On pattern analysis in the non-convex case. Kybernetes 19, 26–33.

Cuevas, A., Fraiman, R., 1993. Patern analysis via nonparametric density estimation. Unpublished manuscript, Departamento de Mathemáticas, Universidad Autónoma de Madrid.

Dahlhaus, 1988. Empirical spectral processes and their application to time series analysis. Stochastic Process. Appl. 30, 69–83.

Davies, P.L., 1987. Asymptotic behaviour of S-estimates of multivariate location parameters and dispersions matrices. Ann. Statist. 15, 1269–1292.
Dudley, R.M., 1978. Central limit theorems for empirical measures. Ann Probab. 6, 899–929.
Dudley, R.M., 1984. A course on empirical processes. École d'Été de Probatilités de Saint Flour XII-1982. Lecture Notes in Mathematics, vol. 1097. Springer, New York, pp. 1–142.
Einmahl, J.H.J., Mason, D. M., 1992. Generalized quantile processes. Ann. Statist. 20, 1062–1078.
Gänssler, P., 1983. Empirical processes. IMS Lecture Notes – Monograph Series 3.
Goldie, C.M., Greenwood, P.E., 1986. Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. Ann. Probab. 14, 817–839.
Grübel, R., 1988. The length of the shorth. Ann. Statist. 16, 619–628.
Hartigan, J.A., 1987. Estimation of a convex density contour in two dimensions. J. Amer. Statist. Assoc. 82, 267–270.
Kiefer, J. 1970. Deviations between the sample quantile process and the sample DF. In: Madan Lal Puri (ed.), Non-parametric techniques in statistical inference. 299–319. Cambridge Univ. Press.
Koltchinskii, V.I., 1994. Komlos–Major–Tusnady approximations for the general empirical process and Haar expansions of classes of functions. J. Theoret. Probab. 7, 73–118.
Lientz, B.P., 1970. Results on nonparametric model intervals. SIAM J. Appl. Math. 19, 356–366.
Molchanov, I.S., 1993. A limit theorem for solutions of inequalities. Unpublished manuscript.
Messer, K., 1993. Personal communication.
Müller, D.W., Sawitzki, 1987. Using excess mass estimates to investigate the modality of a distribution. Preprint Nr. 398 SFB 123, Universität Heidelberg.
Müller, D.W., Sawitzki, G., 1992. Excess mass estimates and tests of multimodality. J. Amer. Statist. Assoc. 86, 738–746.
Nolan, D., 1989. The excess mass ellipsoid. J. Multivariate Anal. 39, 348–371.
Nolan, D., 1992. Asymptotics for multivariate trimming. Stochastic Process. Appl. 42, 157–169.
Pollard, D., 1984. Convergence of stochastic processes. Springer, New York.
Polonik, W., 1992. The excess mass approach to cluster analysis and related estimation procedures. Dissertation, Universität Heidelberg.
Polonik, W., 1995a. Density estimation under qualitative assumptions in higher dimensions. J. Multivariate Anal. 55, 61–81.
Polonik, W., 1995b. Measuring mass concentrations and estimating density contour clusters – an excess mass approach. Ann. Statist. 23, 855–881.
Rousseeuw, P.J., 1986. Multivariate estimation with high breakdown point. In: Grossmann, W., Pflug, G., Vince, I., Wertz, W. (eds.) Mathematical Statistics and Applications. Reidel, Dordrecht, pp. 283–297.
Sager, T.W., 1978. Estimation of a multivariate mode. Ann. Statist. 6, 802–812.
Sager, T.W., 1979. An iterative method for estimating a multivariate mode and isopleth. J. Amer. Statist. Assoc. 74, 329–339.
Sawitzki, G., 1994. Diagnostic plots for one-dimensional data. In: Dirschedl, P., Ostermann, R. (eds.), Papers collected on the Occasion of the 25th Conf. on Statistical Computing at Schloss Reisensburg. Physica, Heidelberg.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Application to Statistics. Wiley, New York.
Tsybakov, A.B., 1997. On nonparametric estimation of level sets. to appear in Am. Statist., no. 3.