Measuring Mass Concentrations and Estimating Density Contour Clusters-An Excess Mass Approach

Wolfgang Polonik

The Annals of Statistics, Vol. 23, No. 3 (Jun., 1995), 855-881.

Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28199506%2923%3A3%3C855%3AMMCAED%3E2.0.CO%3B2-K

The Annals of Statistics is currently published by Institute of Mathematical Statistics.

The Annals of Statistics
1995, Vol. 23, No. 3, 855–881

MEASURING MASS CONCENTRATIONS AND ESTIMATING DENSITY CONTOUR CLUSTERS—AN EXCESS MASS APPROACH[1]

BY WOLFGANG POLONIK

Universität Heidelberg

By using empirical process theory, the so-called excess mass approach is studied. It can be applied to various statistical problems, especially in higher dimensions, such as testing for multimodality, estimating density contour clusters, estimating nonlinear functionals of a density, density estimation, regression problems and spectral analysis. We mainly consider the problems of testing for multimodality and estimating density contour clusters, but the other problems also are discussed. The excess mass (over $\mathbb{C}$) is defined as a supremum of a certain functional defined on $\mathbb{C}$, where $\mathbb{C}$ is a class of subsets of the d-dimensional Euclidean space. Comparing excess masses over different classes $\mathbb{C}$ yields information about the modality of the underlying probability measure F. This can be used to construct tests for multimodality. If F has a density f, the maximizing sets of the excess mass are level sets or density contour clusters of f, provided they lie in $\mathbb{C}$. The excess mass and the density contour clusters can be estimated from the data. Asymptotic properties of these estimators and of the test statistics are studied for general classes $\mathbb{C}$, including the classes of balls, ellipsoids and convex sets.

1. Introduction. The excess mass approach which will be studied in this paper by means of empirical process theory was first considered independently by Müller and Sawitzki (1987) and Hartigan (1987). It can be applied to various statistical problems, especially in higher dimensions, such as testing for multimodality, estimating density contour clusters, estimating nonlinear functionals of a density, density estimation, regression analysis and spectral analysis. In this paper we mainly concentrate on testing for multimodality and estimating density contour clusters, but applications of the excess mass approach to the other problems are also discussed.

Let F be a distribution on $\mathbf{R}^d$ with Lebesgue density f. Müller and Sawitzki defined the *excess mass functional* as $\lambda \to E(\lambda) = F(C(\lambda)) - \lambda \operatorname{Leb}(C(\lambda))$, $\lambda \geq 0$, where Leb denotes Lebesgue measure and $C(\lambda) = C_f(\lambda) = \{x\colon f(x) \geq \lambda\}$ is the *density contour cluster (of f) at a level λ* (cf. Figure 1, which motivates the name excess mass). Note that Hartigan (1975) used the notion density contour cluster, or λ-cluster, for connected components of

Received April 1993; received July 1994.
[1]Research supported by the Sonderforschungsbereich 123 and the Deutsche Forschungsgemeinschaft.
AMS 1991 *subject classifications.* Primary 62G99; secondary 62H99.
Key words and phrases. Excess mass, density contour cluster, level set estimation, multimodality, empirical process theory, support estimation, convex hull.

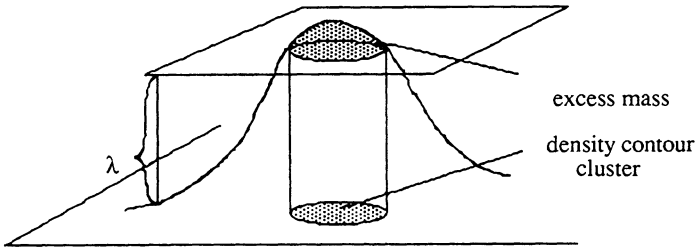

FIG. 1. *Unimodal situation.*

$C(\lambda)$, whereas here the sets $C(\lambda)$ need not be connected.

Let $\mathbb{C}$ denote a class of measurable subsets of $\mathbf{R}^d$ and let $H_\lambda = F - \lambda \,\mathrm{Leb}$. The *excess mass over* $\mathbb{C}$ *at a level* $\lambda \geq 0$ is defined by

$$E_{\mathbb{C}}(\lambda) = \sup\{H_\lambda(C) : C \in \mathbb{C}\}.$$

Every set $\Gamma_{\mathbb{C}}(\lambda) \in \mathbb{C}$ with

$$E_{\mathbb{C}}(\lambda) = H_\lambda(\Gamma_{\mathbb{C}}(\lambda))$$

is called a *generalized* λ*-cluster in* $\mathbb{C}$. Of course, $E_{\mathbb{C}}(\lambda) = E(\lambda)$ if $C(\lambda) \in \mathbb{C}$. In this case $C(\lambda)$ is a generalized λ-cluster. Since $E(\lambda)$ can be rewritten as $E(\lambda) = \sup\{H_\lambda(C),\ C \subset \mathbf{R}^d \text{ measurable}\}$, one has $E_{\mathbb{C}}(\lambda) \leq E(\lambda)$. Replacing F in the definition of $E_{\mathbb{C}}(\lambda)$ by the empirical measure F_n of n i.i.d. observations $X_1, \ldots, X_n$ drawn from F leads to empirical versions, $E_{n,\mathbb{C}}(\lambda)$ and $\Gamma_{n,\mathbb{C}}(\lambda)$, respectively (cf. Section 2).

Differences of excess masses over different classes yield information about modality [Müller and Sawitzki (1987) and Hartigan (1987)]. To see this, assume for the moment that f is a smooth density on the real line with exactly m modes. In this case the density contour clusters $C(\lambda)$ all lie in l_m, $m \in \mathbf{N}$, the class of unions of at most m intervals. Therefore, $E = E_{l_k}$ for all $k \geq m$. Hence, if $\sup_{\lambda \geq 0}(E_{l_k}(\lambda) - E_{l_m}(\lambda))$ is strictly bigger than zero for some $k > m$, then f has more than m modes.

This one-dimensional approach is studied in Müller and Sawitzki. Hartigan considered the two-dimensional case. In our terminology he used the excess mass over the class of closed convex sets in $\mathbf{R}^2$, here denoted by $\mathscr{C}^2$, and compared it with the excess mass over those convex sets lying exterior to $\Gamma_{\mathscr{C}^2}(\lambda)$. In a more parametric setup, Nolan (1991) considered the case $\mathbb{C} = \mathscr{E}^d$, the class of all closed ellipsoids in $\mathbf{R}^d$.

In all these papers it is assumed that the underlying distribution has density contour clusters lying in the class $\mathbb{C}$ under consideration. This assumption, or, for short, the choice of a class $\mathbb{C}$, may be interpreted as the choice of a nonparametric statistical model: the class of all distributions dominated by Lebesgue with $C(\lambda) \in \mathbb{C}$. In contrast to defining models through smoothness assumptions it is possible to model certain *qualitative* aspects, such as modality, of the underlying distribution through appropriate choices

of $\mathbb{C}$. As indicated above, the class l_m corresponds to one-dimensional distributions with at most m modes. Below we also give multivariate analogs.

Density contour clusters contain information about location of mass concentration. If they lie in $\mathbb{C}$, or, in other words, if the chosen model which corresponds to the choice of $\mathbb{C}$ is correct, then they can be estimated by $\Gamma_{n,\mathbb{C}}(\lambda)$. This could also be done by first estimating the density by a kernel estimator, say, and then estimating $C(\lambda)$ by the corresponding density contour cluster of the density estimate. This gives a consistent estimator under appropriate smoothness assumptions. However, this approach does not allow enclosing prior knowledge about the shape of density contour clusters (such as convexity). Furthermore, although one never knows in practice that density contour clusters lie in $\mathbb{C}$, the interpretation of $\Gamma_{n,\mathbb{C}}(\lambda)$ as sets maximizing the excess mass still holds and therefore they might contain useful information even for finite n.

Of course, calculation of $\Gamma_{n,\mathbb{C}}(\lambda)$ for classes $\mathbb{C}$ being not too sparse is time-consuming, especially in higher dimensions. In the class $\mathscr{C}^2$, Hartigan gave an algorithm for calculating $\Gamma_{n,\mathbb{C}}(\lambda)$ which he claimed to have complexity $O(n^3)$. Nolan calculated $\Gamma_{n,\mathbb{C}}(\lambda)$ in the class of ellipsoids in $\mathbf{R}^2$. Note that by definition $\Gamma_{n,\mathbb{C}}(\lambda)$ has minimum Lebesgue measure among all sets in $\mathbb{C}$ which contain not less empirical mass than $\Gamma_{n,\mathbb{C}}(\lambda)$ itself. In other words, $\Gamma_{n,\mathbb{C}}(\lambda)$ is a so-called minimum volume set in $\mathbb{C}$ [to the *random* parameter $\alpha = F_n(\Gamma_{n,\mathbb{C}}(\lambda))$]. Hence, algorithms for calculating minimum volume sets can be used to calculate $\Gamma_{n,\mathbb{C}}(\lambda)$. This has been done by Nolan.

Here we do not explicitly specify $\mathbb{C}$. Therefore one has the flexibility to compare different models. All standard classes, such as the classes of balls, ellipsoids or convex sets, are included in our study. The results of the present paper show how the asymptotic behavior of estimators and test statistics under consideration depend on the richness of $\mathbb{C}$ (and on smoothness assumptions on f). The asymptotic results can be used as hints on how to choose appropriate classes for special problems. One will have to balance between flexibility of the model (which means richness of the classes $\mathbb{C}$ under consideration), desirable statistical properties and time needed for calculation.

The paper is organized as follows: $E_{n,\mathbb{C}}(\cdot)$ is studied in Sections 2 and 4. We show that $E_{n,\mathbb{C}}(\cdot)$ is consistent for $E_{\mathbb{C}}(\cdot)$ and prove asymptotic normality. For example, in the case where $C(\lambda) \in \mathbb{C}$ and "f has no flat part," that is, $F\{x\colon f(x) = \lambda\} = 0\ \forall\ \lambda \geq 0$, the limiting process of $n^{1/2}(E_{n,\mathbb{C}}(\cdot) - E_{\mathbb{C}}(\cdot))$ is a Brownian bridge with transformed time scale (Theorem 4.3). The asymptotic behavior of $\Gamma_{n,\mathbb{C}}(\lambda)$ is studied in Section 3. We show consistency [in $L_1(F)$] as an estimator of $\Gamma_{\mathbb{C}}(\lambda)$ (Theorems 3.2 and 3.5) and in the case where $C(\lambda) \in \mathbb{C}$ we also give rates of convergence (Theorems 3.6 and 3.7). As a special case ($\lambda = 0$ and $\mathbb{C} = \mathscr{C}^d$) we obtain rates of convergence for the convex hull of the sample as a by-product (Proposition 3.8). In Section 5 we study $\sup_{\lambda \geq 0}(E_{n,\mathbb{D}}(\lambda) - E_{n,\mathbb{C}}(\lambda))$, where $\mathbb{C}$ and $\mathbb{D}$ are nested classes, as a test statistic for the hypothesis that the density contour clusters lie in the smaller classes $\mathbb{C}$ against the alternative that they lie in $\mathbb{D} \setminus \mathbb{C}$. For special choices of $\mathbb{C}$ and $\mathbb{D}$ this leads to tests for unimodality, as proposed by Müller and

Sawitzki and by Hartigan (see above). Asymptotic distribution of this test statistic is known only for the special case of an underlying uniform distribution (Theorem 5.4). However, under the null hypotheses, we derive rates of convergence for general F (Theorems 5.2 and 5.3) which in some special situations are known to be the exact rates. Applications of the excess mass approach to other statistical problems are discussed in Section 6. Section 7 contains the proofs of all results given in this paper.

We close the introduction by giving some related work from the literature. There exist other nonparametric approaches to measuring mass concentrations and investigating modality of the underlying distribution, which also are based on the idea of comparing the volume of sets with the mass carried by them:

In a fundamental paper Chernoff (1964) considered the midpoint x of an interval with given length l which carries maximal mass along all intervals with the same length l. If l tends to zero and the distribution is dominated by Lebesgue measure, then, in regular cases, x converges to the mode of the density. However, if l is not too small, the midpoint indicates a location around which a nonnegligible portion of the mass is concentrated. Considered as a function of l the maximal mass $\alpha = \alpha(l)$ becomes the well-known concentration function.

Alternatively, one can consider the inverse problem: fix the mass α and ask for the interval with minimal length among all intervals carrying (at least) mass α. Such intervals are called minimal volume intervals or modal intervals [cf. Lientz (1970), Andrews, Bickel, Hampel, Huber, Rodgers and Tukey (1972), Robertson and Cryer (1974) and Grübel (1988)].

In these one-dimensional situations it is a "natural" decision to use intervals. However, strictly speaking, the choice of intervals is natural only if the underlying distribution has a unimodal Lebesgue density, f. For in this case (under some regularity conditions) the density contour clusters of f are intervals and maximize the (theoretical) functions in the procedures given above. This corresponds to the situation "$C(\lambda) \in \mathbb{C}$" in the context of excess masses (see above).

For generalizing the abovementioned procedures to higher dimensions, there is no "natural" choice of a class of sets, even in the unimodal case. One might for example use the classes of all balls, ellipsoids or convex sets. Sager (1979), for example, generalized the method of Robertson and Cryer (1974) by replacing the class of intervals by the class of convex sets. The problem of how to choose an appropriate class $\mathbb{C}$ (especially in higher dimensions) of course also exists in the excess mass approach. However, as mentioned earlier, since the results in the present paper are given for an unspecified class $\mathbb{C}$, they can be used as hints on how to choose $\mathbb{C}$ in a special problem.

2. The empirical excess mass over $\mathbb{C}$. Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random vectors in $\mathbf{R}^d$ with distribution F. In order to obtain an estimator of the excess mass over $\mathbb{C}$, we replace the unknown distribution F by F_n, the empirical distribution of $X_1, \ldots, X_n$. This leads to the *empirical* excess mass

over $\mathbb{C}$, defined by

$$E_{n,\mathbb{C}}(\lambda) := \sup\{F_n(C) - \lambda \operatorname{Leb}(C) : C \in \mathbb{C}\}, \qquad \lambda \geq 0.$$

Let $H_{n,\lambda} = F_n - \lambda \operatorname{Leb}$, $\lambda \geq 0$. A set $\Gamma_{n,\mathbb{C}}(\lambda) \in \mathbb{C}$ such that

$$E_{n,\mathbb{C}}(\lambda) = H_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda))$$

is called an *empirical generalized λ-cluster*.

Since the "excess" $E_{n,\mathbb{C}}(\lambda)$ [and also $E_{\mathbb{C}}(\lambda)$] should be nonnegative we always assume that $\varnothing \in \mathbb{C}$. In the following proposition some elementary properties of $E_{n,\mathbb{C}}$ are summarized.

PROPOSITION 2.1. *Let $\varnothing \in \mathbb{C}$. Then we have the following*:

(i) $0 \leq E_{n,\mathbb{C}}(\lambda) \leq 1$ *for all* $\lambda \geq 0$;
(ii) $E_{n,\mathbb{C}}(\lambda)$ *is monotone decreasing and convex in* $[0, \infty)$;
(iii) $\lambda \to E_{n,\mathbb{C}}(\lambda)$ *is piecewise linear with at most* $n + 1$ *changes of slope.*

For every distribution F, properties (i) and (ii) also hold for $E_{\mathbb{C}}(\cdot)$.

Consistency. First note that there exist situations where $E_{n,\mathbb{C}}(\lambda)$ contains no information about the underlying distribution [cf. Hartigan (1987)]. Define $A = \{X_1, \ldots, X_n\}$. Then $F_n(A) = 1$ and $\operatorname{Leb}(A) = 0$. Hence, if $A \in \mathbb{C}$, then $E_{n,\mathbb{C}}(\lambda) = 1$ for all $\lambda \geq 0$, independent of F. Therefore $E_{n,\mathbb{C}}$ is in general not a consistent estimator of $E_{\mathbb{C}}$. Consistency Lemma 2.2 given below shows that $E_{n,\mathbb{C}}$ is consistent if $\mathbb{C}$ is a *Glivenko–Cantelli class for F*.

Let (Ω, P) denote the underlying probability space. To avoid measurability considerations, we define for any function $f\colon \Omega \to \mathbf{R}$ the *measurable cover function* f^* as the smallest measurable function from Ω to $\mathbf{R}$ lying everywhere above f [see Dudley (1984)]. Furthermore, let P^* denote *outer probability*. Given a class $\mathbb{C}$ denote

$$\|F_n - F\|_{\mathbb{C}} = \sup\{|(F_n - F)(C)| : C \in \mathbb{C}\}.$$

DEFINITION. A *Glivenko–Cantelli class* (GC-class) for a distribution F, or, for short, a *GC*(F)-class, is a class $\mathbb{C}$ of measurable sets such that, with probability 1,

$$\|F_n - F\|_{\mathbb{C}}^* \to 0 \quad \text{as } n \to \infty.$$

LEMMA 2.2 (Consistency). *For any class $\mathbb{C}$ we have*

$$\sup_{\lambda \geq 0} |E_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)| \leq \|F_n - F\|_{\mathbb{C}}.$$

Hence, if $\mathbb{C}$ is a GC-class for F, then, with probability 1,

$$\sup_{\lambda \geq 0} |E_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)|^* \to 0 \quad \text{as } n \to \infty.$$

It is well known that Vapnik–Cervonenkis (VC) classes are GC-classes for *all* F if in addition they satisfy some measurability assumptions. Examples for such classes are the class of all d-dimensional closed balls $\mathscr{B}^d$ and the class of all d-dimensional closed ellipsoids $\mathscr{E}^d$. There also exist interesting classes which are GC-classes (for certain F) but not VC-classes, as, for example, the class of all closed convex sets in $\mathbb{R}^d$, $d \geq 2$, denoted by $\mathscr{C}^d$. These are GC-classes for all distributions F which have a bounded Lebesgue density [see Eddy and Hartigan (1977) for a characterization of the GC-property of $\mathbb{C}^d$].

As mentioned in the Introduction, the choice of $\mathbb{C}$ is identified with the choice of a statistical model, which consists of those distributions with density contour clusters lying in $\mathbb{C}$. In order to model multimodality, we make the following construction: given a class $\mathbb{C}$ of closed subsets of $\mathbf{R}^d$, let $\mathbb{C}_k$, $k \in \mathbf{N}$, denote the class of sets which can be written as a union of k (possibly empty) sets in $\mathbb{C}$, and let

$$\mathbb{N}_{m,k}(\mathbb{C}) := \left\{ \bigcup_{j=1}^{m} \left(C_j \setminus \check{D}_j\right), C_j \in \mathbb{C}, D_j \in \mathbb{C}_k, j = 1, \ldots, m \right\}, \qquad m, k \in \mathbf{N},$$

where $\check{D}_j$ denotes the open kernel of D_j. Note that the sets in $\mathbb{N}_{m,k}(\mathbb{C})$ are closed by definition and that $\mathbb{C}_m \subset \mathbb{N}_{m,k}(\mathbb{C})$ $\forall$ $m, k \geq 1$. The classes $\mathbb{N}_{m,k}(\mathscr{C}^d)$ seem to be appropriate to model, for example, an underlying mixture of normal distributions (cf. Figures 2 and 3).

The classes $\mathbb{N}_{m,k}(\mathbb{C})$ are special cases of GC-classes which we call *k-constructible* [Alexander (1984) used this terminology]: a class $\mathbb{C}$ in a measurable space $(\mathscr{X}, \mathscr{A})$ is called *k-constructible from a GC-class* $\mathbb{D}$ if there exists a function φ from $\mathbb{D}^k$ to $\mathscr{A}$, constructed from the set operations $\cup$, $\cap$ and ^{c}, such that $\mathbb{C} \subset \varphi(\mathbb{D}^k)$. For example, the class $\mathbb{C} \setminus \mathbb{C} = \{C \setminus D\colon C, D \in \mathbb{C}\}$ is 2-constructible from $\mathbb{C}$. More generally, the classes $\mathbb{N}_{m,k}(\mathbb{C})$ are $m(k+1)$-constructible from $\mathbb{C}$.

If $\mathbb{C}$ is a VC-class, then classes which are k-constructible from $\mathbb{C}$ also are VC-classes, that is, the VC-property of $\mathbb{C}$ carries over to the classes $\mathbb{N}_{m,k}(\mathbb{C})$.

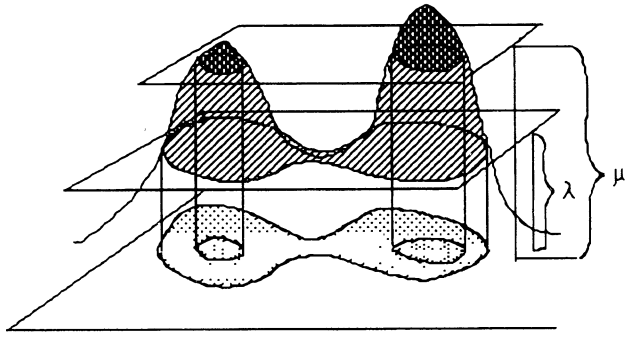

FIG. 2. *Bimodal situation: at the level* μ *the density contour cluster is a union of two disjoint sets. At the level* λ *it is a connected, nonconvex set.*

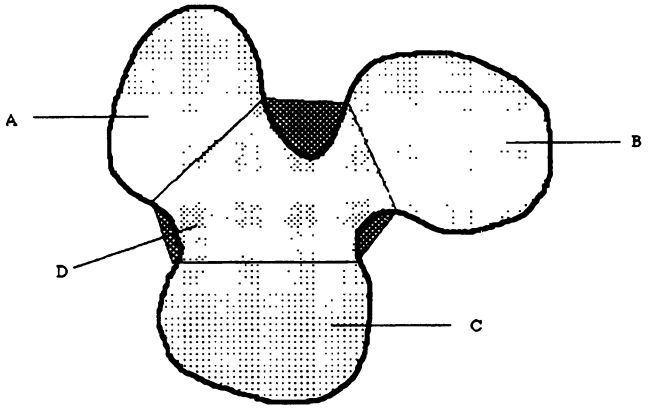

FIG. 3. *A density contour cluster which can be written as a union of the sets A, B, C and D: the sets A, B and C are convex, and D can be obtained by the convex hull of D minus the union of the three dark shadowed sets; therefore this density contour cluster belongs to the class* $\mathbb{N}_{4,3}(\mathscr{E}^2)$.

This is well known [Dudley (1978)]. The analogous property also holds for GC-classes [e.g., see Pollard (1984), Theorem 21 and its proof]. Hence we have the following corollary.

COROLLARY 2.3. *Let* $\mathbb{C}$ *be a GC-class for F. Then for every* $m, k \in \mathbb{N}$ *we have, with probability* 1,

$$\sup_{\lambda \geq 0} |E_{n, \mathbb{N}_{m,k}(\mathbb{C})}(\lambda) - E_{\mathbb{N}_{m,k}(\mathbb{C})}(\lambda)|^* \to 0 \quad \text{as } n \to \infty.$$

3. The empirical generalized λ-clusters. The asymptotic behavior of the empirical generalized λ-clusters $\Gamma_{n,\mathbb{C}}(\lambda)$ is studied in this section. As a measure of distance we use the pseudometric

$$d_F(C, D) := F(C \triangle D), \qquad C, D \in \mathbb{C},$$

where $\triangle$ denotes symmetric difference. Empirical generalized λ-clusters exist for interesting classes $\mathbb{C}$ which consist of *closed* sets, as, for example, for the classes $\mathbb{C} = \mathscr{B}^d$, $\mathscr{E}^d$ or $\mathscr{C}^d$ and for the corresponding classes $\mathbb{N}_{m,k}(\mathbb{C})$ (defined in Section 2). Therefore we assume in all of that what follows that

$$\mathbb{C} \text{ consists of } \textit{closed} \text{ sets.}$$

In addition we assume that

$$\text{Leb}\{\overline{C(\lambda)} \setminus C(\lambda)\} = 0 \quad \text{for all } \lambda \geq 0, \tag{3.1}$$

and only consider

$$\Gamma(\lambda) := \overline{C(\lambda)}$$

the *closure* of the density contour cluster. Because of (3.1) one can still think of $\Gamma(\lambda)$ as the density contour cluster; (3.1) is trivially satisfied for all upper semicontinuous densities, but of course many other densities also have this property.

In the sequel the following assumptions are assumed to hold unless stated otherwise:

GENERAL ASSUMPTIONS.

(A1) For all $\lambda \geq 0$ there exists a generalized and an empirical generalized λ-cluster.

(A2) The underlying distribution F on $\mathbf{R}^d$ has a Lebesgue density f such that $\sup\{(f(x)\} = M < \infty$ and (3.1) holds.

(A3) All classes $\mathbb{C}$ under consideration are GC(F)-classes and consist of closed sets. Furthermore we assume that $\varnothing \in \mathbb{C}$.

As already mentioned, the existence of a set $\Gamma_{\mathbb{C}}(\lambda)$ is guaranteed if $\Gamma(\lambda) \in \mathbb{C}$, but this assumption is not necessary. For every distribution G which has a strictly positive Lebesgue density and every fixed $\lambda \geq 0$, the function $C \to H_\lambda(C)$ is upper semicontinuous on $(\mathbb{C}, d_G)$ (Lemma 7.1). Hence, if the space $(\mathbb{C}, d_G)$ is compact, then a generalized λ-cluster $\Gamma_{\mathbb{C}}(\lambda)$ exists. The latter situation holds, for example, for $\mathbb{C} = \mathscr{B}^d$, $\mathscr{E}^d$ or $\mathscr{C}^d$.

Consistency. First we consider the case where $\Gamma(\lambda)$ is not necessarily assumed to lie in $\mathbb{C}$, or in other words, we consider a situation where the corresponding model (see above) need not be correct.

Note that the sets $\Gamma_{\mathbb{C}}(\lambda)$ and $\Gamma_{n,\mathbb{C}}(\lambda)$ need not be unique. The nonuniqueness of $\Gamma_{n,\mathbb{C}}(\lambda)$ is not crucial, and *the results given below hold for every choice of* $\Gamma_{n,\mathbb{C}}(\lambda)$. This will not be mentioned further in the formulation of the results.

THEOREM 3.2. *Let* $\Lambda \subset [0, \infty)$. *Suppose that the following two conditions hold*:

(i) *For a distribution G with strictly positive Lebesgue density, the space* $(\mathbb{C}, d_G)$ *is quasicompact.*

(ii) *For every* $\lambda \in \Lambda$, *the generalized* λ*-cluster* $\Gamma_{\mathbb{C}}(\lambda)$ *is unique up to F-nullsets.*

Then we have with probability 1 *that*

$$\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))^* \to 0 \quad \text{as } n \to \infty.$$

REMARK. For the class of all closed convex sets with nonempty interior in $\mathbf{R}^2$, the consistency of the empirical generalized λ-cluster (in the Hausdorff metric) was shown by Hartigan (1987) for fixed λ. Müller and Sawitzki (1991a) proved uniform consistency in the one-dimensional case with $\mathbb{C} = l_k$, where they assumed in addition $\Gamma(\lambda) \in \mathbb{C}$. Nolan (1991) considered the case $\mathbb{C} = \mathscr{E}^d$ in a more parametric setup.

There exist interesting situations where the generalized λ-clusters are not unique. Assume, for example, that f is a (smooth) bimodal univariate den-

sity, symmetric around zero, where a mode is defined to be a local maximum of f. Then, for some λ large enough, the density contours cluster is a union of two nonempty intervals, I_1 and I_2, say. If we choose $\mathbb{C}$ as the class of all intervals, then I_1 and I_2 both are generalized λ-clusters.

THEOREM 3.3. *Let $\lambda \geq 0$. Let $\mathscr{M}_{\mathbb{C}}(\lambda) = \{\Gamma \in \mathbb{C}: H_\lambda(\Gamma) = E_{\mathbb{C}}(\lambda)\}$ be the class of all generalized λ-clusters. Suppose that assumption* (i) *of Theorem* 3.2 *holds. Then we have with probability* 1,

$$\inf_{\Gamma \in \mathscr{M}_{\mathbb{C}}(\lambda)} \{d_F(\Gamma_{n,\mathbb{C}}(\lambda), \Gamma)\}^* \to 0 \quad \text{as } n \to \infty.$$

In the sequel we shall assume that $\Gamma(\lambda) \in \mathbb{C}$. In contrast to the more general case considered above, this additional assumption allows us to derive *explicit* upper bounds for $d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))$, which are the key for deriving consistency results and rates of convergence in the case $\Gamma(\lambda) \in \mathbb{C}$.

PROPOSITION 3.4. *Let $\lambda \geq 0$. If $\Gamma(\lambda) \in \mathbb{C}$, then, for every $\eta > 0$,*

$$\begin{aligned} d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \leq\ & F\{x: |f(x) - \lambda| < \eta\} \\ & + \eta^{-1} M\big[(F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) \\ & \qquad - (F_n - F)(\Gamma(\lambda))\big]. \end{aligned} \tag{3.2a}$$

Furthermore, for $\lambda = 0$ we have

$$d_F(\Gamma(0), \Gamma_{n,\mathbb{C}}(0)) \leq (F_n - F)(\Gamma_{n,\mathbb{C}}(0)) - (F_n - F)(\Gamma(0)). \tag{3.2b}$$

The proof of the next theorem follows immediately from (3.2a) together with (A3).

THEOREM 3.5. *Let Λ be a closed subset of the real line such that $\Gamma(\lambda) \in \mathbb{C}$ for all $\lambda \in \Lambda$, and suppose that*

$$\sup_{\lambda \in \Lambda} F\{x: |f(x) - \lambda| < \eta\} \to 0 \quad \text{as } \eta \to 0. \tag{3.3}$$

Then we have, with probability 1, *that*

$$\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))^* \to 0 \quad \text{as } n \to \infty.$$

REMARK. Condition (3.3) says that "F has no flat part in Λ," that is, $F\{x: f(x) = \lambda\} = 0$ for all $\lambda \in \Lambda$. Another equivalent formulation of (3.3) is to say that $\lambda \to \Gamma(\lambda)$ is continuous in Λ for the d_F-pseudometric. This follows from $F\{x: |f(x) - \lambda| < \eta\} = F(\Gamma(\lambda - \eta)) - F(\Gamma(\lambda + \eta)) - F\{x: f(x) = \lambda - \eta\}$.

Rates of convergence. Our two main results on rates of convergence are Theorems 3.6 and 3.7. The first deals with VC-classes $\mathbb{C}$. In the second we also

allow richer classes, where the richness is measured in terms of metric entropy with inclusion of $\mathbb{C}$ with respect to F, which is defined as follows: let

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

Then $\log N_I(\varepsilon, \mathbb{C}, F)$ is called *metric entropy with inclusion* of $\mathbb{C}$ with respect to F.

In the proofs of the theorems given below we shall use results of Alexander (1984) about the behavior of the empirical process. For that reason we shall also use some of his terminology. Alexander studies empirical processes indexed by VC-classes which need to satisfy a certain measurability condition. The corresponding VC-classes are called n-deviation measurable. Here we do not define n-deviation measurability, because all standard VC-classes which we are interested in (the classes of balls, ellipsoids and finite unions and differences of them) satisfy this measurability condition. A class $\mathbb{C}$ is called *(v, m)-constructible VC-class* if $\mathbb{C}$ is m-constructible (as defined in Section 2) 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 v, such that $\mathbb{D}$ "shatters" no set which consists of v points, and $\mathbb{D}$ "shatters" a finite set C iff every $B \subset C$ is of the form $C \cap D$ for some $D \in \mathbb{D}$.

THEOREM 3.6. *Let $\Lambda \subset [0, \infty)$ be closed and let $\mathbb{C}$ be an n-deviation measurable (v, m)-constructible VC-class. Suppose that there exist constants $\gamma, c \geq 0$ such that, for all $\eta > 0$ small enough,*

$$\sup_{\lambda \in \Lambda} F\{x: |f(x) - \lambda| < \eta\} \leq c\eta^{\gamma}. \tag{3.4}$$

If $\Gamma(\lambda) \in \mathbb{C} \ \forall \ \lambda \in \Lambda$, then there exists a constant $K = K(M, c, \gamma, \mathbb{C})$ such that

$$P^*\left(\sup_{\lambda \in \Lambda} d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) > K\left(\frac{n}{\log n}\right)^{-\gamma/(2+\gamma)} \right) \to 0 \quad \text{as } n \to \infty.$$

EXAMPLES. Consider a fixed level $\lambda > 0$. For levels λ where $\|\operatorname{grad} f(x)\|$ is bounded away from zero in a neighborhood of $\{x: f(x) = \lambda\}$ we have $\gamma = 1$. Let f be a smooth unimodal density. Then, for $d = 1$, the density contour clusters are intervals. Hence we choose $\mathbb{C} = l_1$. For $d \geq 2$ we assume the density contour clusters to be balls or ellipsoids, that is, we take $\mathbb{C} = \mathscr{B}^d$ or $\mathscr{E}^d$. For these situations we obtain from Theorem 3.6 that

$$d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_{P^*}\left(n^{-1/3}(\log n)^{1/3}\right).$$

Levels λ where $\gamma < 1$ are called *critical levels*. If, for example, f has a unique

maximum λ_0 at the mode x_0 and behaves like a parabola in a neighborhood of x_0, then it can be shown that

$$F\{x: |f(x) - \lambda_0| < \eta\} = \begin{cases} O(\eta^{1/2}), & \text{for } d = 1, \\ O(\eta), & \text{for } d \geq 2. \end{cases} \tag{3.5}$$

Hence, if f has no additional critical levels $\lambda \neq \lambda_0$, then we have, for $\Lambda = [\delta, \infty)$, $0 < \delta < \lambda_0$, that $\gamma = \frac{1}{2}$ for $d = 1$ and $\gamma = 1$ for $d > 1$. If we consider the same VC-classes as above, that is, $\mathbb{C} = l_1$ for $d = 1$ and $\mathbb{C} = \mathscr{B}^d$ or $\mathscr{E}^d$ for $d \geq 2$, then

$$\sup_{\lambda \geq \delta} d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = \begin{cases} O_{P^*}\left(n^{-1/5}(\log n)^{1/5}\right), & \text{for } d = 1, \\ O_{P^*}\left(n^{-1/3}(\log n)^{1/3}\right), & \text{for } d \geq 2. \end{cases}$$

If we want to include $\delta = 0$, then additional conditions on the tail behavior of f are necessary to control $\sup_{\lambda \geq 0} F\{x: |f(x) - \lambda| < \eta\}$ as $\eta \to 0$.

THEOREM 3.7. *Let $\mathbb{C}$ be such that there exist constants $A, r > 0$ with*

$$\log N_l(\varepsilon, \mathbb{C}, F) \leq A\varepsilon^{-r} \quad \forall\, \epsilon > 0. \tag{3.6}$$

Suppose that there exists a closed subset Λ of $[0, \infty)$ such that $\Gamma(\lambda) \in \mathbb{C}$, for all $\lambda \in \Lambda$, and that (3.4) holds. Then there exist positive constants $L(r) = L(r, A, M)$ such that, with probability tending to 1 as $n \to \infty$

$$\sup_{\lambda \in \Lambda} d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))^* \leq \begin{cases} L(r)n^{-\gamma/(2+1(+r)\gamma)}, & r < 1, \\ L(r)n^{-\gamma/2(\gamma+1)}\log(n), & r = 1, \\ L(r)n^{-\gamma/(\gamma+1)(r+1)}, & r > 1. \end{cases}$$

EXAMPLE. Let $\mathbb{C} = \mathscr{E}^2$ and assume that the sets $\Gamma(\lambda)$, $\lambda \in \Lambda$, all lie in a compact set K. Then we have $r = \frac{1}{2}$ [Dudley (1984)]. Hence, we obtain from Theorem 3.7 for regular situations where $\gamma = 1$ (see above) that $d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_{P^*}(n^{-2/7})$. Hartigan (1987) conjectured that for such cases the rate is $O_{P^*}(n^{-2/7}(\log n)^{2/7})$ in the Hausdorff-distance. If such a compact set K does not exist (e.g., for a distribution with unbounded support and $0 \in \Lambda$), then one in addition needs conditions on the tail behavior of F to ensure that $r = 1$. A sufficient condition is that there exist constants $0 \leq \eta, c, k < \infty$ such that $f(x)\|x\|^\eta \leq c$ for $\|x\| > k$. This is shown in Polonik (1992).

Estimating the support of a density and the case of an underlying uniform distribution. Estimating density contour clusters of a uniform distribution U (for λ not equal to the maximum of the density) means estimating the support of U. Since in this situation the quantity $F\{x: |f(x) - \lambda| < \eta\}$ which appears in (3.2a) is zero for η small enough, we formally have the same basic inequality as in the case of estimating the support of an arbitrary distribution F [cf. (3.2b)]. Therefore we summarize the results concerning both these

cases in Proposition 3.8. The assertion of Proposition 3.8 formally follows from Theorems 3.6 and 3.7, respectively, by taking $\gamma = \infty$.

As mentioned earlier, the support of f, $\operatorname{supp}\{f\}$, is a generalized 0-cluster if it lies in $\mathbb{C}$. For $\mathbb{C} = \mathscr{C}^d$, $d \geq 2$, the convex hull of the sample $X_1, \ldots, X_n$, denoted by conv_n, is an empirical generalized 0-cluster.

PROPOSITION 3.8. *The following results hold with probability tending to* 1 *as* $n \to \infty$.

(i) *Let* $\mathbb{C}$ *be an* n*-deviation measurable* (v, m)*-constructible VC-class and suppose that* $\operatorname{supp}\{f\} \in \mathbb{C}$. *Then there exists a constant* $C = C(v, m)$ *such that*

$$d_F(\operatorname{supp}\{f\}, \Gamma_{n,\mathbb{C}}(0))^* < Cn^{-1}\log(n).$$

(ii) *Let the class* $\mathbb{C}$ *satisfy* (3.6) *with* $r, A > 0$ *and suppose that* $\operatorname{supp}\{f\} \in \mathbb{C}$. *Then there exist constants* $C(r) = C(r, A)$ *such that*

$$d_F(\operatorname{supp}\{f\}, \Gamma_{n,\mathbb{C}}(0))^* < \begin{cases} C(1)n^{-1/2}\log(n), & r = 1, \\ C(r)n^{-1/(1+r)}, & r \neq 1. \end{cases}$$

Hence, if $\mathbb{C} = \mathscr{C}^d$, $d \geq 2$, $m, k \in \mathbf{N}$ *and* $\operatorname{supp}\{f\}$ *is compact, then we have*

$$d_F(\operatorname{supp}\{f\}, \operatorname{conv}_n)^* \leq \begin{cases} C(1)n^{-1/2}\log(n), & d = 3, \\ C\left(\dfrac{d-1}{2}\right)n^{-2/(d+1)}, & d \neq 3. \end{cases}$$

(iii) *Let* U *be a uniform distribution on a bounded set* S, *and denote* $M = 1/\operatorname{Leb}(S)$. *If* $S \in \mathbb{C}$, *then the rates given above also hold for* $\sup_{\lambda < M - \delta} d_U(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))^*$, $\delta > 0$ *arbitrary.*

REMARKS.

(i) Ignoring the log term for $d = 3$ the rates of part (iii) are minimax rates for estimating the support of a uniform distribution [Mammen and Tsybakov (1995)].

(ii) For an underlying uniform distribution with compact convex support in $\mathbb{R}^d$ which has a smooth boundary, it is known that $n^{-2/(d+1)}$ is the exact L_1-rate of the random quantity $d_{\text{Leb}}(\operatorname{supp}\{f\}, \operatorname{conv}_n)$. For $d = 2$ this is a well-known result of Rényi and Sulanke (1964) [cf. Schneider (1988) for a survey of results in this context]. However, also in the case of an unbounded convex support, Proposition 3.6(ii) gives rates of convergence of the convex hull of the sample. We only need to control the metric entropy with bracketing of the corresponding class $\mathbb{C}$. In the example given after Theorem 3.6 we already mentioned that, for $\mathbb{C} = \mathscr{C}^2$, condition (3.6) holds with $r = 1/2$ if a weak condition on the tail behavior is satisfied. Hence, in this case Proposition 3.8(ii) gives $d_F(\operatorname{supp}\{f\}, \operatorname{conv}_n) = O_{P^*}(n^{-2/3})$.

4. The empirical excess mass, revisited. Consistency results and rates of convergence of empirical generalized λ-clusters (derived in the previous

sections) will now be used to study the asymptotic behavior of the standardized empirical excess mass, which is defined as

$$Z_{n,\mathbb{C}}(\lambda) := n^{1/2}\big(E_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)\big).$$

If we (formally) ignore the estimation of $\Gamma_{\mathbb{C}}(\lambda)$ and consider $\tilde{E}_{n,\mathbb{C}}(\lambda) = H_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda))$, then the difference $\tilde{E}_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)$ simply equals the difference $(F_n - F)(\Gamma_{\mathbb{C}}(\lambda))$, which is of the order $O_P(n^{-1/2})$. It will turn out that the random fluctuation which comes in through the estimation of $\Gamma_{\mathbb{C}}(\lambda)$ is asymptotically negligible, so that $n^{1/2}$ is the appropriate normalizing factor. Even in the case where the generalized λ-clusters $\Gamma_{\mathbb{C}}(\lambda)$ are not unique, $Z_{n,\mathbb{C}}(\lambda)$ can be approximated by $F_n - F$ evaluated at the generalized λ-clusters. However, in contrast to the "case of uniqueness," the generalized λ-clusters have to be chosen randomly in $\mathscr{M}_{\mathbb{C}}(\lambda) = \{\Gamma \in \mathbb{C}\colon H_\lambda(\Gamma) = E_{\mathbb{C}}(\lambda)\}$ the class of all generalized λ-clusters (for λ fixed).

We say that the set-indexed empirical process ν_n is *stochastically equicontinuous in the limit* if

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P^*\Big(\sup_{d_F(C,D) < \delta} |\nu_n(C) - \nu_n(D)| > \eta \Big) = 0 \quad \text{for all } \eta > 0.$$

THEOREM 4.1. *Assume that the following two conditions hold*:

(i) *There exists a distribution G which has a strictly positive Lebesgue density such that the space $(\mathbb{C}, d_G)$ is quasicompact.*

(ii) *The empirical process ν_n indexed by $\mathbb{C}$ is stochastically equicontinuous in the limit.*

Let $\lambda \geq 0$ be fixed. Then there exists a random sequence $\{\Gamma_{\mathbb{C}}(\lambda, n),\ n \in \mathbf{N}\} \subset \mathscr{M}_{\mathbb{C}}(\lambda)$ such that

$$\big|Z_{n,\mathbb{C}}(\lambda) - n^{1/2}(F_n - F)\big(\Gamma_{\mathbb{C}}(\lambda, n)\big)\big| = o_{P^*}(1) \quad \text{as } n \to \infty.$$

COROLLARY 4.2. *Assume that conditions* (i) *and* (ii) *of Theorem* 4.1 *hold. Then we have, for every $\lambda \geq 0$ such that $\mathscr{M}_{\mathbb{C}}(\lambda)$ is finite, that*

$$Z_{n,\mathbb{C}}(\lambda) = O_{P^*}(1) \quad \text{as } n \to \infty.$$

The rate is exact if $F(\Gamma) > 0$ for all $\Gamma \in \mathscr{M}_{\mathbb{C}}(\lambda)$.

If the generalized λ-clusters are uniquely determined (up to F-nullsets), then we can prove stronger results. Let $\mathscr{D}(\Lambda)$ denote the space of all real-valued functions on Λ which are continuous from the right and have left limits, equipped with the Skorohod topology.

THEOREM 4.3. *Let $\Lambda \subset [0, \infty)$ be compact. Assume that the generalized λ-clusters are unique up to F-nullsets and that the following conditions hold*:

(i) $\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))^* \to 0$ *with probability* 1 *as* $n \to \infty$ *and* (ii) ν_n *indexed by* $\mathbb{C}$ *is stochastically equicontinuous in the limit. Then*

$$\sup_{\lambda \in \Lambda} |Z_{n,\mathbb{C}}(\lambda) - B_{n,\mathbb{C}}(\lambda)| = o_{P^*}(1) \quad \text{as } n \to \infty,$$

where $B_{n,\mathbb{C}}(\lambda) = n^{1/2}(F_n - F)(\Gamma_{\mathbb{C}}(\lambda))$. *If in addition* (iii) $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ *is continuous in* Λ *for* d_F *and* (iv) $\Gamma_{\mathbb{C}}(\lambda) \subset \Gamma_{\mathbb{C}}(\mu)$ *for* $\mu \leq \lambda$, *then*

$$B_{n,\mathbb{C}}(\lambda) \to B(a_F(\lambda)) \quad \text{in distribution as } n \to \infty$$

in $\mathscr{D}(\Lambda)$, *where* B *denotes a standard Brownian bridge and* $a_F(\lambda) = F(\Gamma_{\mathbb{C}}(\lambda))$.

REMARKS.

(i) If the sets $\Gamma_{\mathbb{C}}(\lambda)$ were not nested, the limiting process would not simply be a Brownian bridge with transformed time axis, but the covariance would be $F(\Gamma_{\mathbb{C}}(\lambda) \cap \Gamma_{\mathbb{C}}(\mu)) - F(\Gamma_{\mathbb{C}}(\lambda))F(\Gamma_{\mathbb{C}}(\mu))$.

(ii) Assumptions (iii) and (iv) are, for example, fulfilled if $\Gamma(\lambda) \in \mathbb{C}\ \forall\, \lambda \geq 0$ and if F has no flat part, that is, (3.3) holds.

Suppose that the assumptions of Theorem 4.3 are satisfied with $\Lambda = \Lambda_0 = [0, \lambda_0]$, $\lambda_0 \geq M$ and $\mathbb{C} = \mathscr{C}^2$. Then we have for every $\varepsilon > 0$ that, as $n \to \infty$,

$$P\left[\sup_{\lambda \in \Lambda_0} \left|Z_{n,\mathbb{N}_{m,k}(\mathbb{C}^2)}(\lambda)\right| \leq \varepsilon\right] \to P\left[\sup_{0 \leq t \leq 1} |B(t)| \leq \varepsilon\right]. \tag{4.1}$$

This leads to confidence bands for $E(\lambda)$. If λ_0 has to be chosen smaller than M, as, for example, in the case of the uniform distribution, or if λ has to be bounded away from zero, then the right-hand side in (4.1) is asymptotically larger than the left-hand side [cf. Müller and Sawitzki (1987) for the one-dimensional case].

5. Tests based on differences of excess masses. In this section the following testing problem is studied: let $\mathbb{C}$ and $\mathbb{D}$ be two classes of measurable subsets of $\mathbb{R}^d$ with $\mathbb{C} \subset \mathbb{D}$, and let Λ be a subset of $[0, \infty)$. We consider the hypothesis that the generalized λ-clusters in $\mathbb{D}$ already lie in the smaller class $\mathbb{C}$, that is, the problem is testing

$$H_0: \Gamma_{\mathbb{D}}(\lambda) \subset \mathbb{C} \quad \text{for all } \lambda \in \Lambda$$

versus

$$H_1: \Gamma_{\mathbb{D}}(\lambda) \subset \mathbb{D} \setminus \mathbb{C} \quad \text{for some } \lambda \in \Lambda,$$

where we assume $\Gamma_{\mathbb{D}}(\lambda)$ to be unique up to F-nullsets for all $\lambda \in \Lambda$. Let $\Delta_n(\mathbb{C}, \mathbb{D}, \lambda) = E_{n,\mathbb{D}}(\lambda) - E_{n,\mathbb{C}}(\lambda)$. As a test statistic for the above testing problem we consider

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = \sup_{\lambda \in \Lambda} \Delta_n(\mathbb{C}, \mathbb{D}, \lambda).$$

This test statistic is a generalization of the test statistics proposed by Müller and Sawitzki (1987) and Hartigan (1987), respectively, for testing the hypoth-

esis of multimodality; $\Delta_n(\mathbb{C}, \mathbb{D}, \lambda)$ is nonnegative for each $\lambda \geq 0$, and large values of this statistic (for some λ) suggest a violation of the hypotheses H_0 (see the Introduction).

If we consider the univariate case and choose $\mathbb{C} = l_1$ and $\mathbb{D} = l_2$, then the above testing problem can be regarded as looking for unimodality versus bimodality (cf. Introduction). For the analogous problem in two dimensions an appropriate choice is $\mathbb{C} = \mathscr{C}^2$ and $\mathbb{D} = \mathbb{N}_{3,2}(C^2)$ (cf. Figure 2). Tests for the hypothesis of "k modes," $k \geq 2$, against the alternative of "m modes," $k < m$, can be constructed analogously. Choosing $\mathbb{C}$ as the class of all balls and $\mathbb{D}$ as the class of all ellipsoids gives a test which may be interpreted as a test for homoscedasticity.

In the important special case where the (closures of the) density contour clusters are assumed to lie in $\mathbb{D}$, the testing problem reduces to

$$\tilde{H}_0: \Gamma(\lambda) \in \mathbb{C} \quad \text{for all } \lambda \in \Lambda$$

versus

$$\tilde{H}_1: \Gamma(\lambda) \in \mathbb{D} \setminus \mathbb{C} \quad \text{for some } \lambda \in \Lambda.$$

Define $\Delta(\mathbb{C}, \mathbb{D}, \lambda) = E_{\mathbb{D}}(\lambda) - E_{\mathbb{C}}(\lambda)$ and $T(\mathbb{C}, \mathbb{D}, \Lambda) = \sup_{\lambda \in \Lambda} \Delta(\mathbb{C}, \mathbb{D}, \lambda)$.

PROPOSITION 5.1. *For every choice of* $\mathbb{C}$ *and* $\mathbb{D}$ *we have*

$$\sup_{\lambda \geq 0} |\Delta_n(\mathbb{C}, \mathbb{D}, \lambda) - \Delta(\mathbb{C}, \mathbb{D}, \lambda)| \leq \|F_n - F\|_{\mathbb{D}} + \|F_n - F\|_{\mathbb{C}}. \tag{5.1}$$

Hence, if $\mathbb{D}$ *is a GC-class for* F, *then for any* $\Lambda \subset [0, \infty)$, *with probability* 1,

$$|T_n(\mathbb{C}, \mathbb{D}, \Lambda) - T(\mathbb{C}, \mathbb{D}, \Lambda)|^* \to 0 \quad \text{as } n \to \infty.$$

If in addition H_0 *holds, then it follows that, with probability* 1,

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda)^* \to 0 \quad \text{as } n \to \infty.$$

Inequality (5.1) immediately follows from Lemma 2.2. If $E_{\mathbb{D}}(\lambda) - E_{\mathbb{C}}(\lambda) > 0$ for some $\lambda \in \Lambda$, then it follows from Proposition 5.1 that the power of a test based on $T_n(\mathbb{C}, \mathbb{D}, \Lambda)$ converges to 1 as n tends to infinity. This is the case if $F(\Gamma_{\mathbb{D}}(\lambda) \Delta \Gamma_{\mathbb{C}}(\lambda)) > 0$ for some $\lambda \in \Lambda$. In general the condition $F(\Gamma_{\mathbb{D}}(\lambda) \Delta \Gamma_{\mathbb{C}}(\lambda)) > 0$ does not follow from $\Gamma_{\mathbb{D}}(\lambda) \neq \Gamma_{\mathbb{C}}(\lambda)$; however, in many standard situations this is the case.

Rates of convergence. The asymptotic distribution of the proposed test statistic is known only for the case of an underlying uniform distribution (cf. Theorem 5.4). However, rates of convergence for the test statistic can be given which give qualitative insight into the behavior of the test statistic under various testing problems, that is, under various classes $\mathbb{C}$ and $\mathbb{D}$ and sets Λ. In general only upper bounds for the rates of convergence of the test statistics are given. At least in some univariate situations these rates are known to be close (up to a log term) to the exact rates.

THEOREM 5.2. *Let $\mathbb{D}$ be an n-deviation measurable (v, m)-constructible VC-class.*

(i) *If* $\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{D}}(\lambda), \Gamma_{n,\mathbb{D}}(\lambda))^* \to 0$ *as* $n \to \infty$, *then, under* H_0,

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = o_{P^*}(n^{-1/2}) \quad \text{as } n \to \infty.$$

(ii) *If* (3.4) *holds with* $\gamma \geq 0$, *then we have under* $\tilde{H}_0$ *that*

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = O_{P^*}\left(\frac{\log n}{n}\right)^{\frac{1+\gamma}{2+\gamma}} \quad \text{as } n \to \infty.$$

EXAMPLES. The case $\Lambda = [0, \infty)$ is interesting here, because the supremum of the density f clearly is unknown. If f is a smooth unimodal density which behaves like a parabola near the mode, then we have $\gamma = \frac{1}{2}$ [cf. (3.5)]. Hence, it follows that

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = O_{P^*}\left(n^{-3/5}(\log n)^{3/5}\right).$$

This rate has already been derived by Müller and Sawitzki (1991a) with the help of the "Hungarian embedding." In higher dimensions, $d \geq 2$, we have in such regular unimodal cases (where the densities behave like a parabola near the mode) that $\gamma = 1$ [cf. (3.5)]. Hence, we have in this case that

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = O_{P^*}\left(n^{-2/3}(\log n)^{2/3}\right).$$

Note that this rate is faster than the rate for the one-dimensional case. This is because the exponent γ is different for $d = 1$ and $d \geq 2$ [cf. (3.5)].

THEOREM 5.3. *Let* $\mathbb{C}$ *satisfy* (3.6) *with* $r > 0$. *For* $\gamma \geq 0$ *let*

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

(i) *If* $\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{D}}(\lambda), \Gamma_{n,\mathbb{D}}(\lambda))^* \to 0$ *as* $n \to \infty$, *then, under* H_0 *as* $n \to \infty$,

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = \begin{cases} o_{P^*}(n^{-1/2}), & r < 1, \\ O_{P^*}(\alpha_n(r)), & r \geq 1. \end{cases}$$

(ii) *Suppose that* (3.4) *holds with* $\gamma \geq 0$. *Then we have under* $\tilde{H}_0$ *that, as* $n \to \infty$,

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = O_{P^*}(\alpha_n(r, \gamma)).$$

EXAMPLE. Again we consider the case $\Lambda = [0, \infty)$. We assume that f has no flat parts, is unimodal and behaves like a parabola near the mode, so that for $d \geq 2$ we have $\gamma = 1$ [cf. (3.5)]. Furthermore we assume that the density

contour clusters are convex, that is, we choose $\mathbb{C} = \mathscr{C}^d$, $d \geq 2$, so that $r = (d-1)/2$. It follows from Theorem 5.3 that

$$T_n(\mathbb{C}, \mathbb{D}, \Lambda) = \begin{cases} O_{P^*}(n^{-4/7}), & d = 2, \\ O_{P^*}(n^{-1/2}\log(n)), & d = 3, \\ O_{P^*}(n^{-1/(d+1)}), & d \geq 4. \end{cases}$$

The next theorem shows [together with (5.1)] that for an underlying uniform distribution $n^{-1/2}$ is the exact rate for the proposed test statistic under H_0 if in addition $\mathbb{D}$ is a Donsker class (with the exception of some degenerate cases, as e.g., $\mathbb{D} = [\varnothing]$). A class $\mathbb{D}$ is called a *Donsker class for F* if the following two conditions (a) and (b) hold: (a) There exists a $\mathbb{D}$-indexed Brownian bridge $B_{\mathbb{D}}$ corresponding to F [or in other words, an F-bridge over $\mathbb{D}$; cf., e.g, Pollard (1984)]; (b) the $\mathbb{D}$-indexed empirical process ν_n can be approximated by versions $B_{n,\mathbb{D}}$ of $B_{\mathbb{D}}$ in the sense that $\|\nu_n - B_{n,\mathbb{D}}\|_{\mathbb{D}} \to 0$ in outer probability. Note that the $\mathbb{D}$-indexed empirical process is stochastically equicontinuous in the limit if $\mathbb{D}$ is a Donsker class. For a Donsker class $\mathbb{D}$ let

$$Z_{\mathbb{D}}(\lambda) := \sup_{D \in \mathbb{D}} \left(B_{\mathbb{D}}(D) - \lambda \operatorname{Leb}(D)\right).$$

THEOREM 5.4. *Let F be a uniform distribution on a bounded set $C_0 \subset \mathbf{R}^d$, and let $\mathbb{D}_0 := \{D \cap C_0, D \in \mathbb{D}\}$. Suppose that $\mathbb{D}$ is a Donsker class for F. Then there exist versions of $B_{n,\mathbb{D}}$ such that, for every interval $\Lambda \subset [0, \infty)$ with $1/\operatorname{Leb}(C_0) \in \operatorname{int}\Lambda$ and every class $\mathbb{C} \subset \mathbb{D}_0$, we have*

$$\left| n^{1/2} T_n(\mathbb{C}, \mathbb{D}, \Lambda) - \sup_{-\infty < \lambda < \infty} \left(Z_{\mathbb{D}_0}(\lambda) - Z_{\mathbb{C}}(\lambda)\right) \right| = o_{P^*}(1) \quad \text{as } n \to \infty.$$

REMARK. Let $J = \int_0^1 (\log N_1(\eta^2, \mathbb{D}, F))^{1/2}\, d\eta$. The finiteness of J is sufficient for the Donsker property [Dudley (1984)].

It seems to be a difficult problem to determine the asymptotic distribution of $T_n(\mathbb{C}, \mathbb{D}, \Lambda)$ for nonuniform distributions. Even for the simplest one-dimensional case ($\mathbb{C} = l_1$ and $\mathbb{D} = l_2$), there exists no solution for this problem until now. However, in order to construct critical values one could try the following strategy. If $\mathbb{D}$ is a Donsker class (with $r < 1$), then it follows from Theorems 5.3 and 5.4 that under $\tilde{H}_0$ the test statistic is asymptotically larger under the uniform distribution than under distributions which have no flat parts. In this situation one could therefore use critical values under a uniform distribution (obtained by Monte Carlo simulation), so that significance of the test could be controlled, at least for large n. In the one-dimensional case simulation studies of Müller and Sawitzki (1987) show that this strategy works well for $n \geq 10$. In higher dimensions such simulation studies have not been done until now.

6. Other applications of the excess mass approach. In this section we briefly indicate how the excess mass approach can be applied to other statistical problems.

Estimating nonlinear functionals of a density. Linear functionals of the excess mass can be used to construct estimates of nonlinear functionals of the density. Here we consider the estimation of $\int f^k(x)\,dx$, $k \in \mathbf{N}$, $k \geq 2$. The key observation for the construction of our estimators for such functionals is the identity

$$k(k-1)\int_0^M \lambda^{k-2} E(\lambda)\,d\lambda = \int f^k(x)\,dx, \tag{6.1}$$

where (as above) $M = \sup f(x)$. Since $E(\lambda) = F(C(\lambda)) - \lambda \operatorname{Leb}(C(\lambda))$ and $\operatorname{Leb}(C(\lambda)) = \int \mathbb{1}_{C(\lambda)}(x)\,dx$, equation (6.1) follows by Fubini's theorem and elementary reformulations [Polonik (1992)]. Hence, an estimator for $\int f^k(x)\,dx$ can be obtained by plugging into the left-hand side of (6.1) the empirical excess mass and a consistent estimator for M. Let

$$T_n(k, \mathbb{C}) = k(k-1)\int_0^{\lambda_{n,\max}} \lambda^{k-2} E_{n,\mathbb{C}}(\lambda)\,d\lambda,$$

where $\lambda_{n,\max}$ is the largest level where the slope of $E_{n,\mathbb{C}}(\lambda)$ changes [cf. Proposition 2.1(iii)]. If $\Gamma(\lambda) \in \mathbb{C}$ for all $\lambda \geq 0$, then $\lambda_{n,\max}$ is a consistent estimator for M. This follows from the fact that $E_{\mathbb{C}}(\lambda) = E(\lambda) = 0$, for $\lambda \geq M$, together with Lemma 2.2. It is not difficult to see that, under the assumptions of Theorem 4.3 as $n \to \infty$,

$$n^{1/2}\left(T_n(k, \mathbb{C}) - \int f^k(x)\,dx\right) \to k(k-1)\int_0^M \lambda^{k-2} B(a_F(\lambda))\,d\lambda \quad \text{in distribution.}$$

Density estimation. If all the density contour clusters of f can be estimated, then of course one can define a density estimate at a point x by summing up all level λ such that x lies in the corresponding estimates of the level sets. In the one-dimensional situation this estimator has been called silhouette by Müller and Sawitzki (1987). They proposed it as a data-analytic tool. It has been studied as a density estimator in Polonik (1993) in the general situation. It is perhaps interesting to note that this estimator can be considered as a generalization of the Grenander estimator of a monotone density [Polonik (1993)].

The regression problem. Consider the following standard regression model $Y_i = r(x_i) + \varepsilon_i$ where $r\colon \mathbf{R}^d \to \mathbf{R}$ is the regression function and ε_i are i.i.d. errors. Suppose that r is integrable such that $R(C) := \int_C r(x)\,dx < \infty$, $C \in \mathbb{C}$. Consider

$$U_{\mathbb{C}}(\lambda) = \sup\{R(C) - \lambda \operatorname{Leb}(C)\colon C \in \mathbb{C}\}, \qquad 0 < \alpha < R. \tag{6.2}$$

The corresponding empirical version is

$$U_{n,\mathbb{C}}(\lambda) = \inf\{R_n(C) - \lambda \operatorname{Leb}(C), C \in \mathbb{C}\}, \qquad 0 < \alpha < R, \tag{6.3}$$

where $R_n(C) = n^{-1}\sum_{i:\, x_i \in C} Y_i$. Minimizing sets in (6.2) are level sets of r if these level sets lie in $\mathbb{C}$. There are practical problems where one is interested in estimating level sets of a regression function, for example, estimating a specific concentration contour of CS-137 [Messer (1993)]. Instead of ν_n, the process

$$e_n(C) = n^{-1/2}\Big(\sum_{i:\, x_i \in C} Y_i - R(C)\Big)$$

appears. Under smoothness assumptions on r one has

$$\begin{aligned} e_n(C) &= n^{-1/2}\Big(\sum_{i:\, x_i \in C} Y_i - \sum_{i:\, x_i \in C} r(x_i)\Big) + o(1) \\ &= n^{-1/2} \sum_{i:\, x_i \in C} \varepsilon_i + o(1). \end{aligned}$$

Set-indexed partial-sum processes of the form $n^{1/2}\sum_{i:x_i \in C}\varepsilon_i$ have been studied in the literature for regular designs [e.g., Bass and Pyke (1984) and Alexander and Pyke (1986); see Goldie and Greenwood (1986) for the case of not necessarily i.i.d. errors]. 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. Instead of the process e_n, the empirical spectral process appears. See Dahlhaus (1988) for results on weak convergence of the empirical spectral process.

7. Proofs.

Proofs of Section 2.

PROOF OF PROPOSITION 2.1. Since $\varnothing \in \mathbb{C}$, (i) follows directly from the definition of the excess mass. The excess mass $E_{n,\mathbb{C}}(\lambda)$ is a supremum over affine linear functions of λ, which either are constant or have a negative slope. Hence $E_{n,\mathbb{C}}(\cdot)$ is monotone decreasing and convex in $[0, \infty)$. Assertion (iii) follows from the fact that the affine linear functions $\lambda \to F_n(C) - \lambda\,\mathrm{Leb}(C)$, $C \in \mathbb{C}$, over which the supremum in the definition of $E_{n,\mathbb{C}}$ is extended have at most $n + 1$ different intercepts. □

PROOF OF CONSISTENCY LEMMA 2.2. Using $H_{n,\lambda} = H_\lambda + (F_n - F)$, we get

$$\begin{aligned} |E_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)| &= \Big|\sup_{C \in \mathbb{C}} H_{\lambda,\lambda}(C) - \sup_{C \in \mathbb{C}} H_\lambda(C)\Big| \\ &\le \sup_{C \in \mathbb{C}} |H_{n,\lambda}(C) - H_\lambda(C)| = \|F_n - F\|_{\mathbb{C}}. \end{aligned}$$

□

Proofs of Section 3. In order to prove Theorem 3.2 we need two lemmas (Lemmas 7.1 and 7.2), which we will prove first.

LEMMA 7.1 (Properties of H_λ).

(i) $\sup_{\lambda \geq 0}\{|H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - H_\lambda(\Gamma_{\mathbb{C}}(\lambda))|^* \to 0$ *with probability* 1 *as* $n \to \infty$.

(ii) *For every distribution* G *which has a strictly positive Lebesgue density, the function* $C \to H_\lambda(C)$, $C \in (\mathbb{C}, d_G)$ *is upper semicontinuous.*

PROOF. (i) From the definition of $\Gamma_{n,\mathbb{C}}(\lambda)$ it follows $H_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) \geq H_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda))$. Together with $H_{n,\lambda} = H_\lambda + F_n - F$ this leads to

$$\begin{aligned} 0 &\leq H_\lambda(\Gamma_{\mathbb{C}}(\lambda)) - H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) \\ &\leq (F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) - (F_n - F)(\Gamma_{\mathbb{C}}(\lambda)), \end{aligned} \tag{7.1}$$

and since $\mathbb{C}$ is a GC-class for F [general assumption (A1)] the assertion follows.

(ii) First note that F is dominated by G [this follows from (A2)]. Therefore it remains to show that $A \to \text{Leb}(A)$ is lower semicontinuous for d_G. In order to see this, let $\{K_n\}$ be a sequence of compact sets in $\mathbf{R}^d$ with $K_n \uparrow \mathbf{R}^d$. Then, clearly,

$$\text{Leb}(A) = \sup_{n \in \mathbf{N}} \text{Leb}(A \cap K_n),$$

and because G has a strictly positive Lebesgue density the functions $A \to \text{Leb}(A \cap K_n)$ are continuous for d_G. Hence, as a supremum over continuous functions, the function $A \to \text{Leb}(A)$ is lower semicontinuous. □

LEMMA 7.2. *Let* $\Lambda \subset [0, \infty)$. *Suppose that conditions* (i) *and* (ii) *of Theorem* 3.2 *are satisfied. Then* $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ *is uniformly continuous in* Λ *for the* d_F*-pseudometric.*

PROOF. Without loss of generality we assume Λ to be compact, because for any $\lambda \geq M$ we have $\text{Leb}(\Gamma_{\mathbb{C}}(\lambda)) = F(\Gamma_{\mathbb{C}}(\lambda)) = 0$. [This follows from the fact that $E_{\mathbb{C}}(\lambda) = 0$ for $\lambda > \max\{f(x)\}$.]

Let $\{\lambda_n, n \in \mathbb{N}\}$ be a sequence in Λ with $\lambda_n \to \lambda_0$, $\lambda_0 \in \Lambda$. Because of the compactness of $\mathbb{C}$ we may assume that $\{\Gamma_{\mathbb{C}}(\lambda_n)\}$ converges to a set $D_0 \in \mathbb{C}$ in the d_G-pseudometric.

First assume $\lambda_0 \in \text{int}\,\Lambda$, the interior of Λ. Since $\lambda_n \to \lambda_0$ we have for a given $\varepsilon > 0$ that $\lambda_0 - \varepsilon \leq \lambda_n \leq \lambda_0 + \varepsilon$ for large enough n. Remember that $H_\lambda(\Gamma_{\mathbb{C}}(\lambda)) = E_{\mathbb{C}}(\lambda)$ and that $E_{\mathbb{C}}(\lambda)$ is monotonically decreasing (Proposition 2.1). Therefore we get, by using the upper semicontinuity of H_λ (Lemma 7.1), that

$$\begin{aligned} H_{\lambda_0+\varepsilon}(\Gamma_{\mathbb{C}}(\lambda_0 + \varepsilon)) &\leq \limsup_n H_{\lambda_n}(\Gamma_{\mathbb{C}}(\lambda_n)) \\ &\leq \limsup_n H_{\lambda_0-\varepsilon}(\Gamma_{\mathbb{C}}(\lambda_n)) \leq H_{\lambda_0-\varepsilon}(D_0). \end{aligned}$$

Letting $\varepsilon \to 0$ we obtain $H_{\lambda_0}(\Gamma_{\mathbb{C}}(\lambda_0)) \le H_{\lambda_0}(D_0)$, and the assertion follows from the assumed uniqueness of the maximum.

If $\lambda_0 \in \overline{\Lambda} \setminus \Lambda$, where $\overline{\Lambda}$ denotes the closure of Λ, then omit the ε on the obvious side in the above inequalities. □

PROOF OF THEOREMS 3.2 AND 3.3. The proof follows the proof of Proposition 2 of Müller and Sawitzki (1991b). First we prove the special case that Λ consists of a single point λ. In this case the proof is very short and shows the main idea.

We may assume that a given realization of the random sequence $\{\Gamma_{n,\mathbb{C}}(\lambda), n \in \mathbb{N}\}$ converges to a set $D_0 \in \mathbb{C}$ in the d_G-pseudometric. Hence it follows from Lemma 7.1(i) and (ii) that, with probability 1,

$$H_\lambda(\Gamma_{\mathbb{C}}(\lambda)) = \limsup_n H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda))^* \le H_\lambda(D_0),$$

and from the assumed uniqueness of the maximum the assertion follows.

Now we consider the general case where $\Lambda \subset [0,\infty)$ is an arbitrary closed set. We need to show that for every sequence $\{\lambda_n \in \Lambda\}$ we have $d_F(\Gamma_{\mathbb{C}}(\lambda_n), \Gamma_{n,\mathbb{C}}(\lambda_n)) \to 0$ with outer probability 1 as $n \to \infty$. It can be assumed that $\lambda_n \to \lambda_0, \lambda_0 \in \Lambda \cup \{\infty\}$.

Since the function $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ is continuous for the d_F-pseudometric (Lemma 7.2), it is enough to show that, with probability 1,

$$F(\Gamma_{n,\mathbb{C}}(\lambda_n) \Delta \Gamma_{\mathbb{C}}(\lambda_0))^* \to 0 \quad \text{as } n \to \infty.$$

For $\lambda_0 < \infty$ the proof is much the same as the proof of the continuity of $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$. The only difference is that here, in addition, the random quantity $H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda))$ comes in. However, $H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda))$ can uniformly be approximated by the nonrandom quantity $H_\lambda(\Gamma_{\mathbb{C}}(\lambda))$ with outer probability 1 [Lemma 7.1(i)].

We may assume that $\{\Gamma_{n,\mathbb{C}}(\lambda_n), n \in \mathbf{N}\}$ converges to a set $D_0 \in \mathbb{C}$ in the d_G-pseudometric. For a given $\varepsilon > 0$ we have $\lambda_0 - \varepsilon \le \lambda_n \le \lambda_0 + \varepsilon$ for large enough n. Hence, by Lemma 7.1(i) and (ii), we get

$$\begin{aligned} H_{\lambda_0+\varepsilon}(\Gamma_{\mathbb{C}}(\lambda_0 + \varepsilon)) &\le \limsup_n H_{\lambda_n}(\Gamma_{\mathbb{C}}(\lambda_n)) \\ &= \limsup_n H_{\lambda_n}(\Gamma_{n,\mathbb{C}}(\lambda_n)) \\ &\le \limsup_n H_{\lambda_0-\varepsilon}(\Gamma_{n,\mathbb{C}}(\lambda_n)) \le H_{\lambda_0-\varepsilon}(D_0), \end{aligned}$$

with outer probability 1. Letting $\varepsilon \to 0$, we obtain $H_{\lambda_0}(\Gamma_{\mathbb{C}}(\lambda_0)) \le H_{\lambda_0}(D_0)$, and as above the assertion follows from the assumed uniqueness of the maximum.

The case $\lambda_0 = \infty$ follows from the fact that f is bounded and that $\Gamma_{\mathbb{C}}(\lambda) = \varnothing$ for λ bigger than the maximum of f.

The proof of Theorem 3.3 is the same as the proof of Theorem 3.2 given for the case $\Lambda = \{\lambda\}$. □

PROOF OF PROPOSITION 3.4. First note that

$$
\begin{aligned}
H_\lambda(\Gamma(\lambda)) - H_\lambda(C) &= \int_{\Gamma(\lambda)} (f(x) - \lambda)\, dx - \int_C (f(x) - \lambda)\, dx \\
&= \int_{\Gamma(\lambda)\setminus C} (f(x) - \lambda)\, dx - \int_{C\setminus\Gamma(\lambda)} (f(x) - \lambda)\, dx \qquad (7.2)\\
&= \int_{\Gamma(\lambda)\Delta C} |f(x) - \lambda|\, dx.
\end{aligned}
$$

Inequality (3.2b) follows directly from identity (7.2). Now we prove (3.2a). To shorten the notation, let $D_{n,\mathbb{C}}(\lambda) = \Gamma_{n,\mathbb{C}}(\lambda)\Delta\Gamma_{\mathbb{C}}(\lambda)$, so that $F(D_{n,\mathbb{C}}(\lambda)) = d_F(\Gamma_{n,\mathbb{C}}(\lambda), \Gamma_{\mathbb{C}}(\lambda))$. We write $F(D_{n,\mathbb{C}}(\lambda))$ as a sum of two terms:

$$
\begin{aligned}
F(D_{n,\mathbb{C}}(\lambda)) &= F(D_{n,\mathbb{C}}(\lambda) \cap \{x\colon |f(x) - \lambda| < \eta\}) \\
&\quad + F(D_{n,\mathbb{C}}(\lambda) \cap \{x\colon |f(x) - \lambda| \geq \eta\}).
\end{aligned}
$$

The first term on the right-hand side is dominated by $F\{x\colon |f(x) - \lambda| < \eta\}$. As for the second term, (7.1) says that

$$
H_\lambda(\Gamma(\lambda)) - H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) \leq (F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) - (F_n - F)(\Gamma(\lambda)).
$$

Thus, because of $f \leq M$, (3.2a) follows from

$$
\begin{aligned}
H_\lambda(\Gamma(\lambda)) - H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) &= \int_{D_{n,\mathbb{C}}(\lambda)} |f(x) - \lambda|\, dx \\
&\geq \eta\, \mathrm{Leb}(D_{n,\mathbb{C}}(\lambda) \cap \{x\colon |f(x) - \lambda| \geq \eta\}). \qquad \square
\end{aligned}
$$

PROOFS OF THEOREMS 3.6 AND 3.7. Let $\{\delta_n\}$ and $\{\eta_n\}$ be sequences of positive real numbers, and define

$$
\begin{aligned}
B_n = \big\{&\exists\, C, D \in \mathbb{C} \text{ such that } d_F(C, D) > \delta_n \text{ and} \\
&d_F(C, D) \leq c\eta_n^\gamma + \eta_n^{-1} M\big[(F_n - F)(D) - (F_n - F)(C)\big]\big\}.
\end{aligned}
$$

Then it easily follows from (3.2a) that for all sequences $\{\delta_n\}$ and $\{\eta_n\}$ we have

$$
P^*\Big[\sup_{\lambda\in\Lambda} d_F(\Gamma_{n,\mathbb{C}}(\lambda), \Gamma(\lambda)) > \delta_n\Big] \leq P^*(B_n).
$$

Hence, we shall look for the "smallest" sequence $\{\delta_n\}$ such that $P^*(B_n) \to 0$. We have

$$
\begin{aligned}
&P^*(B_n) \\
&\leq P^*\left(\sup_{d_F(C,D)>\delta_n} \left|\frac{\big(c\eta_n^\gamma + \eta_n^{-1} M\big[(F_n - F)(D) - (F_n - F)(C)\big]\big)}{d_F(C, D)}\right| > 1\right) \\
&\leq P^*\left(\sup_{d_F(C,D)>\delta_n} \left|\frac{M\big[(F_n - F)(D) - (F_n - F)(C)\big]}{\eta_n d_F(C, D)}\right| > \frac{1}{2}\right) \\
&\quad + P^*\left(\sup_{d_F(C,D)>\delta_n} \left|\frac{c\eta_n^\gamma}{d_F(C, D)}\right| > \frac{1}{2}\right) \\
&= \mathrm{I} + \mathrm{II}.
\end{aligned}
$$

If we choose $\eta_n = \delta_n^{1/\gamma}/2c$, then II equals zero, and it remains to determine $\{\delta_n\}$ such that (with this choice of η_n) I tends to zero as $n \to \infty$. Note that

$$\{d_F(C, D) > \delta_n\} = \bigcup_{k=0}^{k_n} \{2^k\delta_n < d_F(C, D) \le 2^{k+1}\delta_n\},$$

where k_n is chosen as the smallest integer such that $2^{k_n+1}\delta_n \ge 1$. Hence it follows that

$$\begin{aligned} \mathrm{I} &\le \sum_{k=0}^{k_n} P^*\left(\sup_{d_F(C, D) \le 2^{k+1}\delta_n} |(F_n - F)(D) - (F_n - F)(C)| > \frac{2^k\delta_n^{(1+\gamma)/\gamma}}{4Mc} \right) \\ &\le \sum_{k=0}^{k_n} P^*\left(\sup_{A \in (\mathbb{C} \setminus \mathbb{C})_{n,k}} |\nu_n(A)| > \frac{2^{k+1}n^{1/2}\delta_n^{(1+\gamma)/\gamma}}{16Mc} \right) \\ &= \sum_{k=0}^{k_n} p_{n,k}, \end{aligned}$$

where we define $(\mathbb{C} \setminus \mathbb{C})_{n,k} = \{C \setminus D,\ C, D \in \mathcal{C},\ F(C \setminus D) < 2^{k+1}\delta_n\}$. Now we seek conditions to ensure that the last sum converges to zero (as $n \to \infty$). The probabilities $p_{n,k}$ are exactly of the form which is considered in Alexander (1984). For VC-classes $\mathcal{C}$ he derives exponential inequalities of the form

$$P^*\Big(\sup_{A \in \mathcal{C}} |\nu_n(A)| > N \Big) \le 16 \exp(-(1 - \varepsilon)\Psi(N, n, \alpha)),$$

where $\alpha \ge \sup_{A \in \mathcal{C}}\{F(A)(1 - F(A))\}$ and Ψ has to lie in a certain class of functions, including $\Psi(N, n, \alpha) = \Psi_2(N, n, \alpha) = N^2/2\alpha(1 + N/3n^{1/2}\alpha)$, which corresponds to "Bernstein's inequality" (which holds for a single A). Now we fix $\mathcal{C} = (\mathbb{C} \setminus \mathbb{C})_{n,k}$, $\varepsilon = \frac{1}{2}$ and $\Psi = \Psi_2$. For any k and n we have

$$2^{k+1}\delta_n > \sup_{A \in (\mathbb{C} \setminus \mathbb{C})_n} F(A) \ge \sup_{A \in (\mathbb{C} \setminus \mathbb{C})_n} \{F(A)(1 - F(A))\}.$$

Therefore we take $\alpha = 2^{k+1}\delta_n$, and by definition of $p_{n,k}$ the quantity N corresponds to $2^{k+1}n^{1/2}\delta_n^{(1+\gamma)/\gamma}/16Mc$.

Now we split the proof. First we consider the situation of Theorem 3.6. It is easy to see that with $\delta_n = (n/\log n)^{\gamma/(2+\gamma)}$ conditions (2.20), (2.22) and (2.23) of Alexander are fulfilled. Hence, Theorem 2.8 of Alexander (1984) gives the following bound for $p_{n,k}$:

$$p_{n,k} \le 16 \exp\{-2^k n \delta_n^{(2+\gamma)/\gamma}/16^2c^2M^2(1 + \delta_n^{1/\gamma}/48Mc)\},$$

such that for large enough n,

$$\sum_{k=0}^{k_n} p_{n,k} \le 16 \sum_{k=1}^{k_n} \exp\{-2^k \log n/2\, 16^2c^2M^2\},$$

which converges to zero as $n \to \infty$. This proves Theorem 3.6. To finish the proof of Theorem 3.7, we use Corollary 2.4 of Alexander (1984). It is easy to check that if we choose δ_n as the asserted rates of Theorem 3.7, then

condition (2.7) of Alexander is fulfilled. The rest of the proof is the same as above in the situation of Theorem 3.6. □

PROOF OF PROPOSITION 3.8. The idea of the proof is exactly the same as for the proofs of Theorems 3.6 and 3.7 given above. The only difference is that here we use inequality (3.2b) instead of (3.2a). Note that for the uniform distribution $\sup_{\lambda < M-\delta} F\{x\colon |f(x) - \lambda| < \eta\} = 0$ for $\eta < \delta$. Hence (3.2a) reduces to (3.2b) with an additional multiplicative constant $M\eta^{-1}$. Choose $\eta = \eta_0$, and for a sequence $\{\delta_n\}$ of positive real numbers we define

$$B_n = \{\exists\, C, D \in \mathbb{C} \text{ such that } d_F(C, D) > \delta_n \text{ and } d_F(C, D) \le M\eta_0^{-1}[(F_n - F)(D) - (F_n - F)(C)]\},$$

where formally $\eta_0 = 1$ for the case of support estimation [parts (i) and (ii)]. With this definition of B_n the proof works as the proofs of Theorems 3.6 and 3.7. One has to show that

$$\sum_{k=0}^{k_n} P^*\Big(\sup_{A \in (\mathbb{C}\setminus\mathbb{C})_n} |\nu_n(A)| > 2^{k+1} n^{1/2} \delta_n \eta_0 / 8M\Big) \to 0 \quad \text{as } n \to \infty$$

if we choose δ_n as the rates asserted in Proposition 3.8. Here again one can use results of Alexander (1984) in exactly the same way as in the proofs of Theorems 3.6 and 3.7. □

Proofs of Section 4.

PROOF OF THEOREM 4.1. Remember that for every $\lambda \ge 0$ we have

$$E_{n,\mathbb{C}}(\lambda) = H_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) = F_n(\Gamma_{n,\mathbb{C}}(\lambda)) - \lambda \operatorname{Leb}(\Gamma_{n,\mathbb{C}}(\lambda)).$$

From Theorem 3.3 we obtain that for every choice of the empirical generalized λ-clusters $\Gamma_{n,\mathbb{C}}(\lambda)$ there exists a sequence $\{\Gamma_{\mathbb{C}}(\lambda, n),\ n \in \mathbf{N}\} \subset M_{\mathbb{C}}(\lambda)$ such that $d_F(\Gamma_{n,\mathbb{C}}(\lambda), \Gamma_{\mathbb{C}}(\lambda, n))^* \to 0$ with probability 1.

Since every set $\Gamma_{\mathbb{C}}(\lambda, n)$ is a generalized λ-cluster, it follows, with

$$\tilde{E}_{n,\mathbb{C}}(\lambda) := H_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda, n)) = F_n(\Gamma_{\mathbb{C}}(\lambda, n)) - \lambda \operatorname{Leb}(\Gamma_{\mathbb{C}}(\lambda, n)),$$

that

$$n^{1/2}\big(\tilde{E}_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)\big) = n^{1/2}(F_n - F)(\Gamma_{\mathbb{C}}(\lambda, n)).$$

It remains to show that $|n^{1/2}(E_{n,\mathbb{C}}(\lambda) - \tilde{E}_{n,\mathbb{C}}(\lambda))| = o_{P^*}(1)$ as $n \to \infty$. We have

$$\begin{aligned} 0 &\le E_{n,\mathbb{C}}(\lambda) - \tilde{E}_{n,\mathbb{C}}(\lambda) \\ &= H_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - H_\lambda(\Gamma_{\mathbb{C}}(\lambda, n)) + (F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) - (F_n - F)(\Gamma_{\mathbb{C}}(\lambda, n)) \\ &\le (F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) - (F_n - F)(\Gamma_{\mathbb{C}}(\lambda, n)). \end{aligned}$$

Since $d_F(\Gamma_{n,\mathbb{C}}(\lambda), \Gamma_{\mathbb{C}}(\lambda, n))^* \to 0$ with probability 1, the assertion follows from the equicontinuity of the empirical process indexed by $\mathbb{C}$. □

PROOF OF THEOREM 4.3. As in the proof of Theorem 4.1 it follows, with $\tilde{E}_{n,\mathbb{C}}(\lambda) = H_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda))$, that

$$|Z_{n,\mathbb{C}}(\lambda) - B_{n,\mathbb{C}}(\lambda)| = |n^{1/2}\big(E_{n,\mathbb{C}}(\lambda) - \tilde{E}_{n,\mathbb{C}}(\lambda)\big)|$$
$$\leq (F_n - F)(\Gamma_{n,\mathbb{C}}(\lambda)) - (F_n - F)(\Gamma_{\mathbb{C}}(\lambda)).$$

The first assertion follows from the stochastic equicontinuity of ν_n.

The process $B_{n,\mathbb{C}}(\cdot)$ is a random element in $\mathscr{D}(\Lambda)$ because the sets $\Gamma(\lambda)$ are closed. The convergence of the finite-dimensional distributions follows immediately from the multidimensional central limit theorem. The tightness follows from the continuity of $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ together with the asymptotic stochastic equicontinuity of ν_n indexed by $\mathbb{C}$. □

Proofs of Section 5.

PROOF OF THEOREMS 5.2 AND 5.3. First we prove that under $\tilde{H}_0$ we have

$$(7.3)\quad E_{n,\mathbb{D}}(\lambda) - E_{n,\mathbb{C}}(\lambda) \leq (F_n - F)(\Gamma_{n,\mathbb{D}}(\lambda)) + (F_n - F)(\Gamma_{\mathbb{D}}(\lambda)).$$

Since under H_0 every set $\Gamma_{\mathbb{D}}(\lambda)$ is a generalized λ-cluster for $\mathbb{C}$ and $\mathbb{D}$, we have

$$E_{n,\mathbb{D}}(\lambda) - H_\lambda(\Gamma_{\mathbb{D}}(\lambda)) = H_\lambda(\Gamma_{n,\mathbb{D}}(\lambda)) - H_\lambda(\Gamma_{\mathbb{D}}(\lambda)) + (F_n - F)(\Gamma_{n,\mathbb{D}}(\lambda))$$
$$\leq (F_n - F)(\Gamma_{n,\mathbb{D}}(\lambda))$$

and

$$E_{n,\mathbb{C}}(\lambda) - H_\lambda(\Gamma_{\mathbb{D}}(\lambda)) = H_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) - H_\lambda(\Gamma_{\mathbb{D}}(\lambda))$$
$$\geq H_{n,\lambda}(\Gamma_{\mathbb{D}}(\lambda)) - H_\lambda(\Gamma_{\mathbb{D}}(\lambda)) = (F_n - F)(\Gamma_{\mathbb{D}}(\lambda)),$$

and (7.3) follows. Because of (7.3) we have, for any real numbers β_n and δ_n, that

$$\Big\{n^{1/2} \sup_{\lambda\in\Lambda} \Delta_n(\mathbb{C},\mathbb{D},\lambda) > \beta_n\Big\} \cap \Big\{\sup_{\lambda\in\Lambda} d_F(\Gamma(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \leq \delta_n\Big\}$$
$$\subset \Big\{\sup_{A\in(\mathbb{C}\setminus\mathbb{C})_n} |\nu_n(A)| > 2\beta_n\Big\},$$

where $(\mathbb{C}\setminus\mathbb{C})_n = \{C\setminus D,\ C,\ D\in\mathbb{C},\ F(C\setminus D) < \delta_n\}$. Choose δ_n as the rates given in Theorems 3.6 and 3.7, respectively. Then, as in the proofs of Theorems 3.6 and 3.7 the assertions of Theorems 5.2 and 5.3 follow by Theorem 2.8 and Corollary 2.4, respectively, of Alexander (1984). □

PROOF OF THEOREM 5.4. Without loss of generality we assume $\text{Leb}(C_0) = 1$, so that $F(D) = \text{Leb}(D)$ for all $D \in \mathbb{D}_0$. We have, for any class $\mathbb{C}$, that

$$n^{1/2}E_{n,\mathbb{C}}(\lambda) = \sup_{C\in\mathbb{C}} \big(\nu_n(C) - n^{1/2}(\lambda - 1)\text{Leb}(C)\big).$$

Define $Z_{n,\mathbb{C}}(\lambda) = \sup_{C \in \mathbb{C}}(B_{\mathbb{C}}(C) - n^{1/2}(\lambda - 1)\mathrm{Leb}(C))$. Then we have

$$\sup_{\lambda \in \Lambda} |n^{1/2}\big(E_{n,\mathbb{D}_0}(\lambda) - Z_{n,\mathbb{D}_0}(\lambda)\big)| \leq \sup_{D \in \mathbb{D}_0} |\nu_n(D) - B_{\mathbb{D}_0}(D)|,$$

and it follows that

$$\begin{aligned}
&\Big| n^{1/2} T_n(\mathbb{C}, \mathbb{D}_0, \Lambda) - \sup_{\lambda \in \Lambda}\big(Z_{n,\mathbb{D}_0}(\lambda) - Z_{n,\mathbb{C}}(\lambda)\big)\Big| \\
&\quad = \Big| n^{1/2} \sup_{\lambda \in \Lambda}\big(E_{n,\mathbb{D}_0}(\lambda) - E_{n,\mathbb{C}}(\lambda)\big) - \sup_{\lambda \in \Lambda}\big(Z_{n,\mathbb{D}_0}(\lambda) - Z_{n,\mathbb{C}}(\lambda)\big)\Big| \\
&\quad \leq \sup_{D \in \mathbb{D}_0} |\nu_n(D) - B_{\mathbb{D}_0}(D)| + \sup_{C \in \mathbb{C}} |\nu_n(C) - B_{\mathbb{C}}(C)|.
\end{aligned}$$

The assertion now follows, by a continuity argument, from the identity

$$\sup_{\lambda \in \Lambda}\big(Z_{n,\mathbb{D}_0}(\lambda) - Z_{n,\mathbb{C}}(\lambda)\big) = \sup_{\lambda \in \Lambda_n}\big(Z_{\mathbb{D}_0}(\lambda) - Z_{\mathbb{C}}(\lambda)\big),$$

where $\Lambda_n = [n^{1/2}(\lambda_0 - 1), n^{1/2}(\lambda_1 - 1)]$ and $\Lambda = [\lambda_0, \lambda_1]$. □

Acknowledgments. I would like to thank my supervisor Professor D. W. Müller for drawing my attention to the excess mass approach and for his interest and support during the writing of my thesis. Furthermore, I thank the statistics group of the Universität Heidelberg, in particular W. Ehm and E. Mammen, for valuable discussions, hints and remarks concerning the subject.

REFERENCES

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

ALEXANDER, K. S. and 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. and TUKEY, J. W. (1972). *Robust Estimation of Location: Survey and Advances*. Princeton Univ. Press.

BASS, R. F. and 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.

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

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

DAHLHAUS, R. (1988). Empirical spectral processes and their application to time series analysis. *Stochastic Process. Appl.* **30** 69–83.

DEHARDT, J. (1971). Generalizations of the Glivenko–Canelli theorem. *Ann. Math. Statist.* **42** 2050–2055.

DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory* **10** 227–236.

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

DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilités de Saint Flour XII. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

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

EDDY, W. F. and HARTIGAN, J. A. (1977). Uniform convergence of the empirical distribution function over convex sets. *Ann. Statist.* **5** 370–374.

GOLDIE, C. M and 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. (1975). *Clustering Algorithms*. Wiley, New York.

HARTIGAN, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270.

LIENTZ, B. P. (1970). Results on nonparametric modal intervals. *SIAM J. Appl. Math.* **19** 356–366.

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

MESSER, K. (1993). A fast and easy smoothing algorithm with an application to environmental monitoring data. Talk given at the Oberwolfach meeting on Curves, Images and Massive Computation.

MÜLLER, D. W. (1992). The excess mass approach in statistics. Beiträge zur Statistik Nr. 3, Univ. Heidelberg.

MÜLLER, D. W. and SAWITZKI, G. (1987). Using excess mass estimates to investigate the modality of a distribution. Preprint No. 398, SFB 123, Univ. Heidelberg.

MÜLLER, D. W. and SAWITZKI, G. (1991a). Excess mass estimates and tests of multimodality. *J. Amer. Statist. Assoc.* **86** 738–746.

MÜLLER, D. W. and SAWITZKI, G. (1991b). Excess mass estimates and tests of multimodality. Preprint No. 632, SFB 123, Univ. Heidelberg.

NOLAN, D. (1991). The excess-mass ellipsoid. *J. Mutivariate Anal.* **39** 348–371.

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, Univ. Heidelberg.

POLONIK, W. (1993). Density estimation under qualitative assumptions in higher dimensions. Unpublished manuscript.

RAO, R. R. (1962). Relation between weak and uniform convergence of measures and applications. *Ann. Math. Statist.* **33** 659–680.

RÉNYI, A. and SULANKE, R. (1964). Über die konvexe Hülle von n zufällig gewählten Punkten II. *Z. Wahrsch. Verw. Gebiete* **3** 138–147.

ROBERTSON, T. J. and CRYER, J. D. (1974). An iterative procedure for estimating the mode. *J. Amer. Statist. Assoc.* **69** 1012–1016.

SAGER, T. W. (1979). An iterative method for estimating a multivariate mode and isopleth. *J. Amer. Statist. Assoc.* **74** 329–339.

SCHNEIDER, R. (1988). Random approximation of convex sets. *Journal of Microscopy* **151** 211–227.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720