

CONCENTRATION AND GOODNESS-OF-FIT IN HIGHER DIMENSIONS: (ASYMPTOTICALLY) DISTRIBUTION-FREE METHODS¹

BY WOLFGANG POLONIK

University of California, Davis

A novel approach for constructing goodness-of-fit techniques in arbitrary (finite) dimensions is presented. Testing problems are considered as well as the construction of diagnostic plots. The approach is based on some new notions of mass concentration, and in fact, our basic testing problems are formulated as problems of “goodness-of-concentration.” It is this connection to concentration of measure that makes the approach conceptually simple. The presented test statistics are continuous functionals of certain processes which behave like the standard one-dimensional uniform empirical process. Hence, the test statistics behave like classical test statistics for goodness-of-fit. In particular, for simple hypotheses they are asymptotically distribution free with well-known asymptotic distribution. The simple technical idea behind the approach may be called a generalized quantile transformation, where the role of one-dimensional quantiles in classical situations is taken over by so-called minimum volume sets.

1. Introduction. Many goodness-of-fit techniques in one-dimensional situations are based on distances between the empirical and the hypothetical distribution functions. To this category belong well-known tests like the Kolmogoroff–Smirnov test and Cramér–von Mises test and several modifications thereof as well as graphical techniques like P-P-plots and Q-Q-plots. These techniques, however, lack easy extensions to higher dimensions. As is well known, one reason for this is that the probability integral or quantile transformation does not work in higher dimensions, so that the corresponding straightforward extensions of the above mentioned tests to higher dimensions are no longer distribution free.

A successful approach for extending one-dimensional goodness-of-fit techniques to higher dimensions has been developed by Khmaladze (1981, 1988, 1993), who used a martingale approach to construct transformations of the empirical process which led to asymptotically distribution free test statistics for goodness-of-fit. Foutz (1980) studied another approach based on empirical probability measures. As a test statistic Foutz proposed the maximal difference of the empirical and hypothetical distribution over a certain data-depen-

Received March 1997; revised November 1998.

¹Supported in part by the Deutsche Forschungsgemeinschaft and by European Union Human Capital and Mobility Program ERB CHRX-CT 940693.

AMS 1991 *subject classifications*. 62G10, 62G30, 62-09.

Key words and phrases. Diagnostic plots, empirical process theory, generalized quantile transformation, Kolmogoroff–Smirnov test, minimum volume sets.

dent class of sets, which are constructed out of so-called statistically equivalent blocks. Another related work is Bickel and Breiman (1983).

The approach studied in the present paper also is based on the idea of comparing empirical and hypothetical probability measures over a certain class of sets. The underlying idea, which formally can be understood as a generalized quantile transformation, also allows the construction of diagnostic plots which can be considered as generalized P-P-plots. Moreover, the method presented can be used to construct techniques for inspecting mass concentration of the underlying distribution. It is this connection to concentration that helps to get a more intuitive understanding of the techniques presented.

Let us first recall the classical (one-dimensional) quantile transformation. For the moment let F denote a one-dimensional distribution function, and F_n the empirical distribution function based on n i.i.d. observations drawn from F . Assume that F is continuous. Then it is a classical result that the distribution of

$$KS_n = \sup_{x \in \mathbf{R}} |F_n(x) - F(x)| = \sup_{\alpha \in [0, 1]} |F_n(F^{-1}(\alpha)) - \alpha|$$

does not depend on F , where F^{-1} denotes the (generalized) inverse of F . Moreover, the asymptotic distribution of $\sqrt{n}KS_n$ is well known, namely, the distribution of $\sup_{\alpha \in [0, 1]} |B(\alpha)|$ for a standard Brownian bridge B . These facts, among others, make statistics based on the empirical process in one dimension very attractive. A generalization to higher dimensions is the following. Identify the quantiles $F^{-1}(\alpha)$ with the interval $(-\infty, F^{-1}(\alpha)]$ and replace this family of sets by a family $\{C(\alpha), \alpha \in [0, 1]\}$ with (i) $C(\alpha) \subset C(\beta)$ for $0 \leq \alpha \leq \beta \leq 1$ and (ii) $F(C(\alpha)) = \alpha \forall \alpha \in [0, 1]$. Then, clearly, the statistic $\sup_{\alpha \in [0, 1]} |(F_n - F)(C(\alpha))|$ has exactly the same distribution as KS_n . This is easy to see and is part of the statistical folklore. However, it is worth pointing out that this property holds regardless of the dimension of the underlying space. The important question now is, which family of sets to choose in order to get a “good” test statistic for our goodness-of-fit problem? Here we propose to use minimum volume sets.

Let X_1, X_2, \dots be i.i.d. observations drawn from a probability measure F on a measurable space $(\mathcal{X}, \mathcal{A})$, and let $\mathcal{C} \subset \mathcal{A}$. Let further F_n denote the empirical measure putting mass $1/n$ on each of the first n observations.

DEFINITION 1.1. Let ν be a real-valued function on \mathcal{C} and let $\alpha \in [0, 1]$. Any set $C_{F, \mathcal{C}}(\alpha) \in \mathcal{C}$ such that

$$C_{F, \mathcal{C}}(\alpha) \in \operatorname{argmin}_{C \in \mathcal{C}} \{ \nu(C) : F(C) \geq \alpha \}$$

is called a minimum volume set (MV-set) in \mathcal{C} for F at level α with respect to ν . Instead of $C_{F, \mathcal{C}}(\alpha)$ we write $C_{n, \mathcal{C}}(\alpha)$ and call them empirical MV-sets.

Obviously, quantiles are MV-sets in $\mathcal{C} = \{(-\infty, x], x \in \mathbf{R}\}$ with respect to $\nu((-\infty, x]) = x$, if $x \in \mathbf{R}$ is identified with the interval $(-\infty, x]$. Below we

mainly think of ν as a dominating measure for F , like the Lebesgue measure on \mathbf{R}^d , which is denoted by Leb .

The prototype of our test statistic for the hypothesis $H_0: F = G$, where G is a fixed known probability measure, is of the form

$$(1) \quad T_{n,\mathcal{E}} = \sup_{\alpha \in [0,1]} \left[|(F_n - G)(C_{G,\mathcal{E}}(\alpha))| + |(F_n - G)(C_{n,\mathcal{E}}(\alpha))| \right].$$

The crucial role of the choice of \mathcal{E} will of course be discussed in detail below. If the sup were taken over the first summand only, then $T_{n,\mathcal{E}}$ would be a direct generalization of the Kolmogoroff–Smirnov test statistic. Such types of statistics will also be considered below. However, the second summand in general is necessary (and sufficient) in order to get a test for goodness-of-fit, consistent against all (or at least a large class) of alternatives. Lemma 1.2 motivates the construction of $T_{n,\mathcal{E}}$ using level sets as special MV-sets. A more general version of this lemma is given in Section 2 below.

For a (nonnegative) real-valued function f let

$$\Gamma_f(\lambda) = \{x: f(x) \geq \lambda\}$$

denote the *level set* of f at level $\lambda \geq 0$. Note that if $\nu = \text{Leb}$ and f has no flat parts, that is, $\nu\{x: f(x) = \lambda\} = 0, \forall \lambda > 0$, then the class of level sets, appropriately reparametrized, satisfies conditions (i) and (ii) above.

LEMMA 1.2. *Let ν be a measure on $(\mathcal{X}, \mathcal{A})$. Suppose that F and G are probability measures on $(\mathcal{X}, \mathcal{A})$ with ν -densities f and g , respectively. The following two statements are equivalent:*

- (a) $F = G$;
- (b) $F(\Gamma_g(\lambda)) = G(\Gamma_g(\lambda))$ and $G(\Gamma_f(\lambda)) = F(\Gamma_f(\lambda))$ for almost all $\lambda > 0$.

That level sets of probability densities are MV-sets with respect to the underlying dominating measure is a basic fact and can be seen easily by observing that $F(C) - \lambda\nu(C) = \int_C f(x) - \lambda d\nu(x)$ is maximized over all measurable sets by $\Gamma_f(\lambda)$. In particular, if $\Gamma_f(\lambda) \in \mathcal{E}$ then it is a MV-set in \mathcal{E} at the level $\alpha = F(\Gamma_f(\lambda))$. Hence, the above test statistic $T_{n,\mathcal{E}}$ can be understood as an estimator for the quantity

$$(2) \quad \sup_{\lambda > 0} \left[|F(\Gamma_g(\lambda)) - G(\Gamma_g(\lambda))| + |G(\Gamma_f(\lambda)) - F(\Gamma_f(\lambda))| \right],$$

which under the conditions of Lemma 1.2 equals zero iff H_0 holds. Using only one of the summands in (2) is not enough to get such an equivalence, as the example $F = \mathcal{N}(0, 1), G = \sqrt{\chi^2}$, shows, where $G(\Gamma_f(\lambda)) = F(\Gamma_f(\lambda)), \forall \lambda \geq 0$.

We want to stress here that using $T_{n,\mathcal{E}}$, or its modifications introduced below, makes sense in more general situations than the one given in Lemma 1.2. For instance, the existence of a density is not necessary at all for defining MV-sets and hence for defining $T_{n,\mathcal{E}}$, and MV-sets have a clear interpretation as regions carrying a high mass concentration even if they are not level sets.

The crucial point, of course, is the choice of the class \mathcal{E} . Here again level sets will be used as a guideline.

Note that as in classical one-dimensional situations our test statistic is a continuous functional of processes indexed by a *one-dimensional* parameter in $[0, 1]$. As indicated above, these processes behave asymptotically as the classical uniform empirical process under appropriate assumptions. It follows that for simple hypotheses these test statistics are in every (finite) dimension asymptotically distribution free and that they have the same well-known limiting distribution as the corresponding classical one-dimensional test statistics for goodness-of-fit. Moreover, this structure allows the construction of P-P type plots in higher dimensions. Such plots, which we call C-C-plots, contain information about if and where the two distributions under consideration differ.

In order to formalize the connection to concentration functions, we introduce the notion of \mathcal{E} -concentration in Section 2. The formulation of our basic testing problems using \mathcal{E} -concentration can also be found there, as well as some discussion on the choice of the class \mathcal{E} . Section 3 deals with asymptotic properties of the test statistics. We state consistency results and formulate conditions under which the test statistics have the above mentioned distribution free limit. In Section 4 some modifications of the above test statistic, which have an improved computational performance, are studied. The construction of diagnostic plots (C-C-plots) is considered in Section 5. Section 6 contains some discussion, and the proofs can be found in Section 7.

2. Comparing concentrations. In all of what follows we denote by F, G, H probability measures on $(\mathcal{X}, \mathcal{A})$. For a measure μ on \mathcal{A} and $C, D, \in \mathcal{A}$ we let

$$d_\mu(C, D) = \mu(C\Delta D),$$

where Δ denotes set theoretic symmetric difference. We say that the MV-sets for a distribution F in \mathcal{E} are *uniquely determined* up to ν -nullsets if any two MV-sets in \mathcal{E} at the same level α have d_ν -distance 0.

First we are going to present our basic Lemma 2.1. This lemma also underlies the central notion of \mathcal{E} -concentration defined below. For a distribution F with density f we let $\mathcal{L}_f = \{\Gamma_f(\lambda), \lambda \geq 0\}$ denote the family of all level sets of f .

Consider the following two properties:

$$P_1: F(C_{G, \mathcal{E}}(\alpha)) = G(C_{G, \mathcal{E}}(\alpha)) \text{ and } G(C_{F, \mathcal{E}}(\alpha)) = F(C_{F, \mathcal{E}}(\alpha));$$

$$P_2: d_\nu(C_{F, \mathcal{E}}(\alpha), C_{G, \mathcal{E}}(\alpha)) = 0.$$

LEMMA 2.1. *Suppose that MV-sets in \mathcal{E} for F and G , respectively, are determined uniquely up to ν -nullsets:*

- (a) *For every $\alpha \in [0, 1]$: $P_1 \Rightarrow P_2$.*
- (b) *P_1 holds $\forall \alpha \in [0, 1] \Leftrightarrow P_2$ holds $\forall \alpha \in [0, 1]$.*

Note that Lemma 1.2 does not follow from Lemma 2.1 trivially. In fact, under the assumption $\mathcal{L}_f \cup \mathcal{L}_g \subset \mathcal{E}$ level sets of f and g are MV-sets in \mathcal{E} ; however at different levels α . So that $d_\nu(C_{F,\mathcal{E}}(\alpha), C_{G,\mathcal{E}}(\alpha)) = 0$ for some fixed α does not imply $d_\nu(\Gamma_f(\lambda), \Gamma_g(\lambda)) = 0$ for some λ .

It is interesting to note that P_1 and P_2 are not equivalent for a fixed α in general. As an example just take $\mathcal{E} = \{\emptyset, A, \mathcal{A}\}$, where $A \in \mathcal{A}$ is such that $0 < F(A) < G(A) < 1$. Then it is easy to see that for $\alpha \in (0, F(A)]$ we have $C_{F,\mathcal{E}}(\alpha) = A = C_{G,\mathcal{E}}(\alpha)$, and hence, P_2 holds, but not P_1 .

It follows from Lemma 2.1 together with Lemma 1.2 that if $\mathcal{L}_f \cup \mathcal{L}_g \subset \mathcal{E}$ then “ P_1 holds $\forall \alpha \in [0, 1]$ ” iff “ $F = G$ ” or

$$(3) \quad \sup_{\alpha \in [0, 1]} \left[|F(C_{G,\mathcal{E}}(\alpha)) - G(C_{G,\mathcal{E}}(\alpha))| + |G(C_{F,\mathcal{E}}(\alpha)) - F(C_{F,\mathcal{E}}(\alpha))| \right] = 0 \quad \text{iff } F = G.$$

If we think of MV-sets as (one-dimensional) quantiles, then (3) can be read as $F = G$ iff (i) $F(G^{-1}(\alpha)) = G(G^{-1}(\alpha))$ and (ii) $G(F^{-1}(\alpha)) = F(F^{-1}(\alpha))$, where here F, G denote the corresponding distribution functions. In this case, however, it is actually enough to consider only one of the functions $F(G^{-1})$ or $G(F^{-1})$, since we have $F = G$ iff either (i) or (ii) holds.

Lemma 2.1 also motivates considering the function $(FG^{-1})_{\mathcal{E}}$ from $[0, 1]$ into $[0, 1]$ defined through

$$(FG^{-1})_{\mathcal{E}}(\alpha) = F(C_{G,\mathcal{E}}(\alpha)).$$

$(FG^{-1})_{\mathcal{E}}$ can be used to compare concentrations of F and G . The notation $(FG^{-1})_{\mathcal{E}}$ comes from the fact that in the situation where the one-dimensional quantiles are MV-sets, we have $(FG^{-1})_{\mathcal{E}}(\alpha) = F(G^{-1}(\alpha))$, where on the right-hand side F and G denote distribution functions. Of course we need uniqueness of MV-sets for $(FG^{-1})_{\mathcal{E}}$ to be well defined.

Basic assumptions.

1. ν is a σ -finite measure on $(\mathcal{X}, \mathcal{A})$, and F and G have ν -densities f, g , respectively.
2. MV-sets in \mathcal{E} for F and G are uniquely determined up to ν -null sets.

The basic assumptions can be relaxed at several places, but for clarity of presentation we assume them to hold in all of what follows.

DEFINITION 2.2 (\mathcal{E} -concentration). Given ν we say that $F \in \text{CONC}_{\mathcal{E}}(G)$ iff

$$(FG^{-1})_{\mathcal{E}} = (GG^{-1})_{\mathcal{E}}.$$

We say that “ F and G have the same \mathcal{E} -concentration”, or $F =_{\mathcal{E}} G$, iff

$$G \in \text{CONC}_{\mathcal{E}}(F) \quad \text{and} \quad F \in \text{CONC}_{\mathcal{E}}(G).$$

We first list some properties of \mathcal{E} -concentration.

- (A) Using Lemma 2.1(a) it is easy to see that “ $=_{\mathcal{E}}$ ” defines an equivalence relation on the class of probability distributions with uniquely determined MV-sets in \mathcal{E} .
- (B) In view of Lemma 2.1, $F =_{\mathcal{E}} G$ means that F and G have the same MV-sets in \mathcal{E} up to ν -nullsets.
- (C) It is easy to construct examples where $F =_{\mathcal{E}} G$, but $F \neq G$. However, Lemma 1.2 says that $F =_{\mathcal{E}} G$ iff $F = G$, provided $\mathcal{L}_f \cup \mathcal{L}_g \subset \mathcal{E}$.
- (D) In general, $F \in \text{CONC}_{\mathcal{E}}(G)$ does not define an equivalence relations on the same large class of distributions as “ $=_{\mathcal{E}}$ ” does [cf. (A)]. For instance, we have $\mathcal{N}(0, 1) \in \text{CONC}_{\mathcal{E}}(\sqrt{\chi^2})$, but $\sqrt{\chi^2} \notin \text{CONC}_{\mathcal{E}}(\mathcal{N}(0, 1))$, if \mathcal{E} is the class of all intervals on the real line and $\nu = \text{Leb}$ (cf. Introduction). However, we have the following lemma.

LEMMA 2.3. *Let F and G be such that $\mathcal{L}_f \subset \mathcal{L}_g \subset \mathcal{E}$. Then*

$$F \in \text{CONC}_{\mathcal{E}}(G) \Rightarrow F = G.$$

- (E) Suppose that $\mathcal{L}_g \subset \mathcal{E}$ and that g has a global maximum at 0, say. If f and g are continuous at 0 and $f(0) \neq g(0)$ then it is easy to see that $F \notin \text{CONC}_{\mathcal{E}}(G)$, because F and G cannot give the same mass to level sets of g close to the mode.

Using the above definitions we now formulate the two testing problems that we are interested in. The first problem is

$$H_0^{(1)}: F \in \text{CONC}_{\mathcal{E}}(G) \quad \text{versus} \quad H_1^{(1)}: F \notin \text{CONC}_{\mathcal{E}}(G).$$

This testing problem in general is a problem for testing for concentration rather than for goodness-of-fit. Our second problem is

$$H_0^{(2)}: F =_{\mathcal{E}} G \quad \text{versus} \quad H_1^{(2)}: F \neq_{\mathcal{E}} G.$$

As a test statistic for the problem $H^{(1)}$, we propose

$$(4) \quad S_{n, \mathcal{E}} = \sup_{\alpha \in [0, 1]} |(F_n - G)(C_{G, \mathcal{E}}(\alpha))|,$$

and for the problem $H^{(2)}$ we propose to use $T_{n, \mathcal{E}}$. Comparing $S_{n, \mathcal{E}}$ and $T_{n, \mathcal{E}}$ one can roughly say that $S_{n, \mathcal{E}}$ only “looks in one direction” (from F to G), whereas $T_{n, \mathcal{E}}$ looks in both (from F to G and from G to F).

For the rest of this section we discuss some aspects concerning the choice of \mathcal{E} .

\mathcal{E} -concentration and order restrictions. For the interpretation of \mathcal{E} -concentration we often used the assumption $\mathcal{L}_f \subset \mathcal{E}$. This assumption defines a shape restriction on f . If, in addition, \mathcal{E} is a σ -lattice (closed under countable unions and intersections), then we know that the assumption $\mathcal{L}_f \subset \mathcal{E}$ is equivalent to an order restriction on f , which means that f is monotone with respect to some (pseudo) order determined through \mathcal{E} , where \mathcal{E} is the class of

so-called upper sets for the order restriction [e.g., Robertson, Wright and Dykstra (1988)]. Suppose we know that f satisfies an order restriction and we choose \mathcal{E} as the corresponding class of upper sets. Then Lemma 1.2 says that there is no distribution G different from F with ν -density g satisfying this order restriction and having the same \mathcal{E} -concentration as F .

Sparse classes \mathcal{E} and the curse of dimensionality. The above indicates that it is reasonable to choose \mathcal{E} such that the level sets of the density of the hypothetical distribution lie in \mathcal{E} . This means, in order to have a rich model, we should choose \mathcal{E} as a rich class. In contrast, from a mathematical point of view sparse classes are better, because they lead to better asymptotic properties (see Section 3). An example for a (very) sparse class \mathcal{E} is the class \mathcal{E}_0 of balls in \mathbf{R}^d with fixed midpoint 0, say. If the density of g is strictly positive, then MV-sets in \mathcal{E}_0 for G are uniquely determined, and every set in \mathcal{E}_0 is a MV-set for G , such that, in particular (i) and (ii) from the introduction are fulfilled. It follows from Theorem 3.5 below that under $H_0^{(1)}$ the statistic $S_{n,\mathcal{E}}$ is distribution free with the same distribution for every dimension! Almost trivially, our two basic testing problems coincide in such simple situations. In the above situation we are just testing whether F and G put the same probability mass on each of the balls with midpoint 0. Of course, this has nothing to do with MV-sets directly. However, in practice such types of testing problems might be helpful, especially in higher dimensions. It should be possible to extract more useful information out of the data by choosing several such sparse testing classes than by using, let's say, the half-space distance. Deviations from the hypothetical distribution can be interpreted more easily, especially because one also has available corresponding diagnostic plots (see Section 5).

3. Asymptotics of the test statistics. In the following we assume that all considered random quantities are measurable. Otherwise, we would have to work with "outer" and "inner" probabilities and integrals, respectively [cf. van der Vaart and Wellner (1996)]. Let us recall here that the basic assumptions formulated in Section 2 are assumed to hold everywhere.

Consistency: Some terminology from empirical process theory is used in the sequel. A class \mathcal{E} is called the *Glivenko–Cantelli class* for a distribution F , or *GC(F)-class*, iff $\sup_{C \in \mathcal{E}} |F_n(C) - F(C)| \rightarrow 0$ a.s. as $n \rightarrow \infty$. Well-known examples for *GC(F)-classes* are the classes of intervals, balls and ellipsoids in \mathbf{R}^d . They possess the *GC(F)-property* for all F . Convex sets in \mathbf{R}^d also form *GC(F)-classes*, however, not for all F . A sufficient condition for F to ensure the *GC(F)-property* of the convex sets is that F has a bounded Lebesgue density [cf. Eddy and Hartigan (1977) for necessary and sufficient conditions].

THEOREM 3.1. *Suppose that the class of all MV-sets in \mathcal{E} for G is a GC(F)-class. Then we have*

$$S_{n,\mathcal{E}} \rightarrow 0 \text{ a.s. as } n \rightarrow \infty \quad \Leftrightarrow \quad F \in \text{CONC}_{\mathcal{E}}(G).$$

The assumption that the MV-sets for G form a $GC(F)$ -class obviously holds if \mathcal{E} itself is a $GC(F)$ -class. However, even if \mathcal{E} is too rich to be a $GC(F)$ -class the MV-sets in \mathcal{E} might form such a class, as for instance if $\nu = \text{Leb}$ on \mathbf{R}^d , the density of g has no flat part, and $\mathcal{L}_g \subset \mathcal{E}$. In this case the class of MV-sets in \mathcal{E} equals the class \mathcal{L}_g which is a totally ordered class of sets with respect to inclusion. Such classes are so-called Vapnik–Cervonenkis classes and hence they also have the $GC(F)$ -property. [For a definition of Vapnik–Cervonenkis classes, see, e.g., van der Vaart and Wellner (1996).]

THEOREM 3.2. *Suppose that \mathcal{E} is a $GC(F)$ -class. Then we have the following:*

- (a) $F = G \Rightarrow T_{n,\mathcal{E}} \rightarrow 0$ a.s. as $n \rightarrow \infty$.
- (b) Let $\alpha \in [0, 1]$ be fixed. If
- (5) $\sup_{\alpha \in [0, 1]} |(F_n F_n^{-1})_{\mathcal{E}}(\alpha) - (FF^{-1})_{\mathcal{E}}(\alpha)| \rightarrow 0$ a.s. as $n \rightarrow \infty$, and
- (6) $d_G(C_{F,\mathcal{E}}(\alpha), C_{n,\mathcal{E}}(\alpha)) \rightarrow 0$ a.s. as $n \rightarrow \infty$,

then we have as $n \rightarrow \infty$ that

$$T_{n,\mathcal{E}} \geq |F(C_{G,\mathcal{E}}(\alpha)) - G(C_{G,\mathcal{E}}(\alpha))| + |G(C_{F,\mathcal{E}}(\alpha)) - F(C_{F,\mathcal{E}}(\alpha))| + o(1) \quad \text{a.s.}$$

Hence, if (5) and (6) hold for every $\alpha \in [0, 1]$, then

$$F \neq_{\mathcal{E}} G \Rightarrow \exists c > 0 \text{ s.th. } T_{n,\mathcal{E}} \geq c + o(1) \text{ a.s. as } n \rightarrow \infty.$$

- (c) If (5) and (6) hold uniformly in $\alpha \in [0, 1]$, then
- (7) $F =_{\mathcal{E}} G \Leftrightarrow T_{n,\mathcal{E}} \rightarrow 0$ a.s. as $n \rightarrow \infty$.

Assumption (5) is satisfied in many situations, for instance, if $\nu = \text{Leb}$ and \mathcal{E} is the class of closed balls, ellipsoids or convex sets. Consistency of $C_{n,\mathcal{E}}(\alpha)$ needs to be controlled in view of (6). Suppose that \mathcal{X} is a finite set, ν is a counting measure and \mathcal{E} consists of subsets of \mathcal{X} . Then it is easy to see that for every fixed α we have $C_{n,\mathcal{E}}(\alpha) = C_{F,\mathcal{E}}(\alpha)$ for n large enough, such that in this simple case (6) is satisfied trivially. Under some regularity assumptions, consistency of MV-sets has been studied in Polonik (1997). Besides consistency results, also rates of convergence can be found there in the special case where $\mathcal{L}_f \subset \mathcal{E}$. We state two uniform consistency results which can be proved by using ideas from that paper.

LEMMA 3.3. *Let \mathcal{E} be a $GC(F)$ -class and suppose that (5) holds.*

(a) *Suppose that $(FF^{-1})_{\mathcal{E}} = (GG^{-1})_{\mathcal{E}}$ and that (\mathcal{E}, d_G) is quasi-compact. Suppose further that on each compact set, G is equivalent to ν and that the function $\nu(C_{F,\mathcal{E}}(\cdot))$ is continuous on $[0, 1]$. Then*

$$\sup_{\alpha \in [0, 1]} d_G(C_{n,\mathcal{E}}(\alpha), C_{F,\mathcal{E}}(\alpha)) \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

(b) Let $\nu = \text{Leb}$ on \mathbf{R}^d . Suppose that f has no flat parts, and that $\mathcal{L}_f \subset \mathcal{E}$. If in addition g is bounded, then

$$\sup_{\alpha \in [0, 1]} d_G(C_{n, \mathcal{E}}(\alpha), C_{F, \mathcal{E}}(\alpha)) \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

We would like to stress here, that in part (a) we do not use assumptions on level sets at all. In part (b) we only assume that $\mathcal{L}_f \subset \mathcal{E}$, but not that $\mathcal{L}_g \subset \mathcal{E}$. Of course, the assumptions could be weakened if we would look at pointwise consistency of MV-sets only. The condition $(FF^{-1})_{\mathcal{E}} = (GG^{-1})_{\mathcal{E}}$ of part (a) holds automatically if $F =_{\mathcal{E}} G$ (see the proof of Lemma 2.1).

The next theorem, which covers the goodness-of-fit case, follows from Theorem 3.2(b), together with Lemma 3.3(b) and Lemma 1.2.

THEOREM 3.4. *Let $\nu = \text{Leb}$ on \mathbf{R}^d . Suppose that \mathcal{E} is a $GC(F)$ -class and $\mathcal{L}_f \cup \mathcal{L}_g \subset \mathcal{E}$. If (5) holds, f has no flat parts and g is bounded, then we have*

$$F = G \quad \Leftrightarrow \quad T_{n, \mathcal{E}} \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

In particular, this theorem says that if there is a priori information available indicating that the underlying distribution on \mathbf{R}^d is spherically symmetric and unimodal, and hence we choose \mathcal{E} as the class of balls, then $T_{n, \mathcal{E}}$ is a consistent test against all other spherically symmetric, unimodal, distributions. If we choose \mathcal{E} as the class of convex sets in \mathbf{R}^d as another, more general, “model” for unimodality then $T_{n, \mathcal{E}}$ is consistent against all distributions with convex level sets, provided the underlying density is bounded [to ensure the $GC(F)$ -property of the convex sets]. However, the assumption that both F and G have densities with level sets in \mathcal{E} is not necessary. The class of alternatives that can be detected by means of $T_{n, \mathcal{E}}$ actually is much larger than the one given in Theorem 3.4, as follows from Theorem 3.2(b), together with property (E) from Section 2.

Asymptotic distribution under the null hypotheses. The next result about the asymptotic distribution of $S_{n, \mathcal{E}}$ follows directly from the definition of $F \in \text{CONC}_{\mathcal{E}}(G)$ together with the generalized quantile transformation discussed in the Introduction.

THEOREM 3.5. *Suppose that F, G and \mathcal{E} are such that $F \in \text{CONC}_{\mathcal{E}}(G)$. Suppose further that $C_{G, \mathcal{E}}(\alpha) \subset C_{G, \mathcal{E}}(\beta)$ for $0 \leq \alpha \leq \beta \leq 1$, and the mapping $\alpha \rightarrow F(C_{G, \mathcal{E}}(\alpha))$ is one-to-one onto $[0, 1]$. Then the distribution of $S_{n, \mathcal{E}}$ under $H_0^{(1)}$ is exactly the same as the distribution of the (one-dimensional) Kolmogoroff–Smirnov statistic KS_n for an underlying uniform distribution. In particular, as $n \rightarrow \infty$,*

$$\sqrt{n} S_{n, \mathcal{E}} \rightarrow \sup_{\alpha \in [0, 1]} |B(\alpha)| \quad \text{in distribution,}$$

where B denotes a standard Brownian bridge on $[0, 1]$.

The “continuity assumptions” on MV-sets for G used in Theorem 3.5 are satisfied, if, for instance, G has a ν -density g without flat parts and $\mathcal{L}_g \subset \mathcal{E}$. If the underlying distribution has flat parts, then, although the corresponding level sets do not satisfy these “continuity assumptions,” a class \mathcal{E} can in general be chosen such that the corresponding MV-sets do. This is an advantage of using MV-sets instead of level sets in our basic approach.

For the next theorem we need the notion of asymptotic equicontinuity. Let $\{\nu_{n,\mathcal{E}}(C) = \sqrt{n}(F_n - F)(C), C \in \mathcal{E}\}$, denote the \mathcal{E} -indexed empirical process. Then $\nu_{n,\mathcal{E}}$ is called *asymptotically equicontinuous* if for each $\varepsilon, \delta > 0$ there exists an $\eta > 0$ such that

$$\limsup_{n \rightarrow \infty} P \left\{ \sup_{\{C, D \in \mathcal{E}: d_F(C, D) \leq \eta\}} |\nu_{n,\mathcal{E}}(C) - \nu_{n,\mathcal{E}}(D)| \geq \varepsilon \right\} \leq \delta.$$

For many standard classes \mathcal{E} , such as intervals, balls, or ellipsoids in \mathbf{R}^d the empirical process has this property. For further examples see, for instance, Dudley (1984).

THEOREM 3.6. *Suppose that all the assumptions of Theorem 3.5 hold. If in addition:*

- (i) $\sup_{\alpha \in [0,1]} d_F(C_{n,\mathcal{E}}(\alpha), C_{F,\mathcal{E}}(\alpha)) = o_p(1)$ as $n \rightarrow \infty$ and
- (ii) $\nu_{n,\mathcal{E}}$ is asymptotically equicontinuous,

then we have under the hypothesis “ $F = G$ ” that as $n \rightarrow \infty$,

$$\sqrt{n} T_{n,\mathcal{E}} \rightarrow 2 \sup_{\alpha \in [0,1]} |B(\alpha)| \text{ in distribution,}$$

where B denotes a standard Brownian bridge on $[0, 1]$.

Again, the asymptotic distribution of $T_{n,\mathcal{E}}$ is independent of the underlying distribution F , and the same well-known asymptotic distribution appears. Hence, the above results can easily be used to construct asymptotically valid tests for $H^{(1)}$ and $H^{(2)}$, respectively.

4. Modifications of $T_{n,\mathcal{E}}$. For calculating $T_{n,\mathcal{E}}$ it is necessary to calculate all the empirical MV-sets. This is in general not an easy task. Let, for example, \mathcal{E} be the class of convex sets in \mathbf{R}^3 . The calculation of a convex hull of n data points in three dimensions already is a formidable task, and this is just one empirical MV-set in the class of convex sets. Even in the class of ellipsoids, calculation of all MV-sets still takes too much time to really use the corresponding test statistic in practice. From an asymptotical point of view, however, there is no need to use empirical MV-sets. Other consistent estimators for the theoretical MV-sets can be used as well. Statistics of the form

$$\tilde{T}_{n,\mathcal{E}} = \sup_{\alpha \in [0,1]} \left[|(F_n - G)(C_{G,\mathcal{E}}(\alpha))| + |(F_n - G)(\tilde{C}_n(\alpha))| \right]$$

can be used, where $\tilde{C}_n(\alpha)$ denotes *some* estimator of $C_{F, \mathcal{E}}(\alpha)$. If $\tilde{C}_n(\alpha)$ has the same consistency properties as $C_{n, \mathcal{E}}(\alpha)$ as an estimator of $C_{F, \mathcal{E}}(\alpha)$, then $T_{n, \mathcal{E}}$ and $\tilde{T}_{n, \mathcal{E}}$ essentially give equivalent tests. In particular, $\tilde{T}_{n, \mathcal{E}}$ will also be asymptotically distribution free under mild assumptions. We do not formulate explicit results here, since the modifications of the results of Section 3 are more or less obvious.

4.1. *Estimating the shape of MV-sets.* Here we discuss the possibility of estimating the shape of MV-sets first and then calculating the MV-sets in a much smaller, estimated class of sets. For instance, consider ellipsoids in \mathbf{R}^d of the form $C_{\mu, \Sigma}(c) = \{x \in \mathbf{R}^d: (x - \mu)' \Sigma (x - \mu) \leq c\}$ and suppose the underlying distribution has elliptical contours $C_{\mu_0, \Sigma_0}(c)$, $c \geq 0$. Choose $\mathcal{E} = \{C_{\mu, \Sigma}(c), c \geq 0, \mu \in \mathbf{R}^d, \Sigma$ a positive semidefinite $d \times d$ matrix}. Then, in a first step, one can estimate the “true” values μ_0 and Σ_0 by some estimator $\hat{\mu}$ and $\hat{\Sigma}$, respectively. Several estimators are available here. Then, in a second step, the MV-sets are calculated in the estimated class of ellipsoids $\{C_{\hat{\mu}, \hat{\Sigma}}(c), c \geq 0\}$. This second step in fact is a one-dimensional problem and can be solved very fast computationally. This idea of approximating MV-ellipsoids of course is not new [see, e.g., Davies (1987)].

4.2. *Using density estimators.* If the underlying distribution has a density with level sets lying in the class \mathcal{E} under consideration, then these level sets are MV-sets. Hence, one can try to use a density estimator and use level sets of this estimator as estimators for MV-sets. In contrast to the theory above we have to give up the idea of including shape restrictions (by prespecifying a class \mathcal{E}). Moreover, densities with flat parts have to be excluded to get the desired asymptotic distribution (see also the discussion after Theorem 3.5).

Let \hat{f}_n denote a density estimator for f , and let $\hat{\Gamma}_n(\lambda)$ denote the corresponding level sets of \hat{f}_n . We consider statistics of the form

$$\hat{T}_n = \sup_{\lambda \geq 0} \left(|(F_n - G)(\Gamma_g(\lambda))| + |(F_n - G)(\hat{\Gamma}_n(\lambda))| \right).$$

In principle, there is no need for prespecifying a class \mathcal{E} . However, for deriving asymptotic results, a class \mathcal{E} containing all the estimated level sets comes in again.

First we formulate a general consistency result for $\hat{\Gamma}_n(\lambda)$ which says that pointwise consistent density estimators give uniformly consistent estimators for level sets.

PROPOSITION 4.1. *Let \hat{f}_n be any density estimator converging to f in probability a.e. If f is bounded and has no flat parts, then*

$$\sup_{\lambda \geq 0} d_F(\hat{\Gamma}_n(\lambda), \Gamma_f(\lambda)) = o_P(1) \quad \text{as } n \rightarrow \infty.$$

This result can in principle be combined with the results of Section 3 to get consistency and a distributional result for \hat{T}_n . However, this does not always go through without problems, as our first example shows.

4.2.1. *Histogram estimators.* Let us consider histograms based on regular grids. Level sets of such estimators can be handled computationally quite easily in arbitrary dimension, and the corresponding test based on \hat{T}_n is consistent (under mild additional assumptions) if the number of cells of the regular grid does not grow too fast. This can be seen as follows. Let \mathcal{P}_{k_n} denote a regular grid in \mathbf{R}^d where k_n is the number of partitions in each dimension, such that k_n^d is the number of overall cells, and assume that the cells have boundaries parallel to the coordinate axes. If $k_n^d = O(n/\log n)$ then the σ -algebra generated by these cells has the $GC(F)$ -property for any distribution F . This can be proved by using well-known results from empirical process theory [e.g., Alexander (1984), Pollard (1984)]. However, as for the distributional result, it turns out, that the class of possible level sets of histograms based on regular grids is growing too fast in order to get the same asymptotic distribution as for $T_{n,\mathcal{E}}$ [condition (ii) of Theorem 3.6 cannot be verified].

4.2.2. *Kernel estimators.* In contrast to the histograms considered above, level sets of a kernel estimator \hat{f}_n can be forced to lie (asymptotically) in a certain class \mathcal{E} of sets with smooth boundaries. Asymptotic properties of the empirical process indexed by classes of sets depend on the richness of the indexing class, measured, for instance, in terms of so-called metric entropy with inclusion. There exist several results ensuring, for instance, the asymptotic equicontinuity of the empirical process indexed by classes of smooth sets [e.g., Dudley (1974)]. In general, however, the higher the dimension, the stronger are the smoothness assumptions needed. To make all this mathematically precise, lots of technicalities are necessary. We do not formulate such a result here. Everything seems to work under appropriate assumptions.

4.3. *Granulometric smoothing.* Another way of estimating level sets of densities has recently been proposed in Walther (1997). His approach also depends on density estimation. However, in contrast to the above, no assumptions to ensure the convergence of the derivatives of the estimators are necessary. By construction the estimators lie in a class of sets with smooth boundaries. The idea is the following. First, pick out subsets of the observations by means of a density estimator \hat{f}_n , namely, look at the sets $\mathcal{X}_n^+(\lambda) = \{X \in \{X_1, \dots, X_n\} : \hat{f}_n(X) \geq \lambda\}$. Then these sets of points are “smoothed” by means of a smoothing operation called “granulometry” [Matheron (1975)]. The resulting sets $\Gamma_n(\lambda)$ which lie in a class \mathcal{E}_r of sets with smooth boundaries (with r as a smoothness parameter; see below) are the final estimators of the level sets $\Gamma(\lambda)$ and can be used in \hat{T}_n . Empirical processes indexed by this class of sets can be controlled, as follows from results in Walther (1997), such that under appropriate conditions $\hat{T}_{n,\mathcal{E}}$ based on $\Gamma_n(\lambda)$ has the same asymptotically distribution free limit distribution as $T_{n,\mathcal{E}}$. The notion of smoothness used in this context is connected to Blaschkes rolling theorem. An analytical definition is as follows [cf. Theorem 1, Walther (1997)]: $S \in \mathcal{E}_r$ if ∂S is a $(d - 1)$ -dimensional C^1 submanifold in \mathbf{R}^d with outward pointing

unit normal vector $n(s)$ at $s \in \partial S$ satisfying the Lipschitz condition

$$|n(s) - n(t)| \leq \frac{1}{r}|s - t| \quad \text{for all } s, t \in \partial S.$$

The smoothing parameter r has to tend to zero not too fast as the sample size tends to infinity, in order to ensure a good rate of convergence. Such results can be found in Walther (1997). However, these results are not uniform over all $\lambda \geq 0$ as is needed in our context. For achieving this, some additional technical assumptions seem to be necessary.

5. C-C-plots. As already discussed in Section 2, the functions $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ contain information about the concentration of F relative to G and vice versa. We use these functions to construct generalized P-P-plots. This is in analogy to the one-dimensional case where $F(G^{-1})$ and $G(F^{-1})$ are used to construct P-P-plots. Note, however, $F(G^{-1})$ and $G(F^{-1})$ are inverse functions if F and G are continuous. However, for the functions $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ this is only true in very special cases (see below). And in general, it is essential to consider *both* functions.

Motivated by Lemma 2.1 we define a C-C-plot (“C” here stands for concentration) as a plot consisting of the graphs of the two curves

$$(8) \quad \alpha \rightarrow (F_n G^{-1})_{\mathcal{E}}(\alpha) \quad \text{and} \quad \alpha \rightarrow (G F_n^{-1})_{\mathcal{E}}(\alpha)$$

together with the diagonal. Note that in order to calculate the second function in (8) we have to calculate all the empirical MV-sets in \mathcal{E} . This can be quite time consuming. Therefore, analogously to the idea underlying the construction of the modified test statistic $T_{n,\mathcal{E}}$ considered in Section 4, one can use the two curves $\lambda \rightarrow F_n(\Gamma_g(\lambda))$ and $\lambda \rightarrow G(\Gamma_n(\lambda))$ and rescale them. This leads to *modified* C-C-plots based on the functions

$$(9) \quad \alpha \rightarrow F_n(\Gamma_g(\lambda_g(\alpha))) \quad \text{and} \quad \alpha \rightarrow G(\hat{\Gamma}_n(\lambda_{\hat{f}_n}(\alpha))),$$

where \hat{f}_n denotes a density estimator, and where $\lambda_f(\alpha)$ is defined through the equation $F(\Gamma_f(\lambda_f(\alpha))) = \alpha$. Similarly, we define $\lambda_{\hat{f}_n}(\alpha)$. Calculation of $G(\hat{\Gamma}_n(\lambda_{\hat{f}_n}(\alpha)))$ is feasible by means of Monte Carlo integration.

Figure 1(a)–(c) shows such (modified) empirical C-C-plots using a kernel density estimator with a normal kernel. The graphs of the two curves are linearized there. They are simulated with both F and G , a standard normal distribution in different dimensions, respectively. The sample size is $n = 25$, and the bandwidth is chosen by hand. (Our simulation studies indicate that the plots are not very sensitive to the choice of the bandwidth if under-smoothing is avoided.) The computer programs for all the figures in this section are written in Oberon and make use of Voyager [Sawitzki, Diller and Friedrich (1994)].

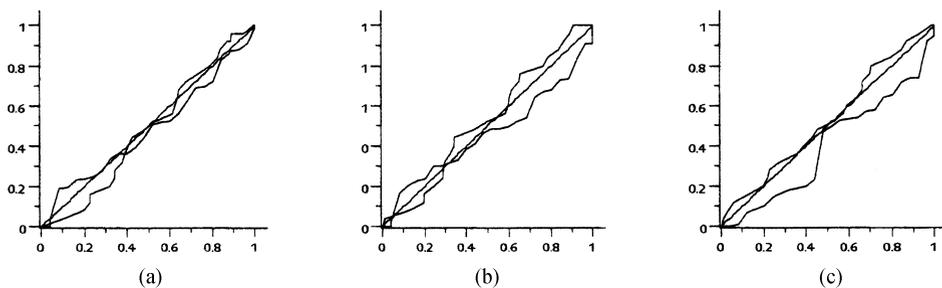


FIG. 1. Modified empirical C-C-plots under $H_0^{(2)}$ based on 25 observations from a standard normal distribution. (a) $d = 1$, (b) $d = 3$, (c) $d = 10$.

Properties and interpretations of C-C-plots. In order to get some feeling about the interpretation of C-C-plots, some properties of the functions $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ are discussed now. Plots of these two functions (together with the diagonal) are called *theoretical C-C-plots*. The essential property of a theoretical C-C-plot is the following: provided F and G both have densities without flat parts and level sets in \mathcal{E} , then:

Both graphs in the theoretical C-C-plot are straight lines through the origin with slope one if and only if $F = G$.

Of course, if we consider modified theoretical C-C-plots (using level sets of densities instead of MV-sets) then the class \mathcal{E} plays no role here.

Deviations from the diagonal near zero, near one, or in the middle, indicate deviations of F and G near the mode(s), in the tails, or in the main body of the distribution, respectively.

For the rest of this section, we only consider modified C-C-plots using level sets of densities in \mathbf{R}^d with $\nu = \text{Leb}$. In the modified theoretical C-C-plots shown in this section for different F and G , the graph of $(FG^{-1})_{\mathcal{E}}(\alpha) = F(\Gamma_g(\lambda_g(\alpha)))$ is plotted as a solid line, and the graph of $(GF^{-1})_{\mathcal{E}}(\alpha) = G(\Gamma_f(\lambda_f(\alpha)))$ as a dashed line. The probability measures F and G are given for each figure.

There is some close connection between C-C-plots and P-P-plots in the following special situation. Suppose that both F and G are distributions on the real line with densities f and g symmetric around the same location parameter, zero, say. Suppose in addition that f and g are unimodal in the sense that all their level sets are intervals with midpoints zero. Then we have $F(\Gamma_g(\lambda_g(\alpha))) = 2F(G^{-1}((1 + \alpha)/2)) - 1$. It follows that $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ are inverse functions. The corresponding theoretical P-P-plot is given by the graph of $\alpha \rightarrow F(G^{-1}(\alpha))$. Under the present assumptions, P-P-plots are symmetric around the diagonal with respect to the point $(1/2, 1/2)$; that is, we have $F(G^{-1}((1 + \alpha)/2)) = 1 - F(G^{-1}((1 - \alpha)/2))$. Hence, in this situation (theoretical) C-C-plots and P-P-plots contain the same information (e.g., Figure 2a, c).

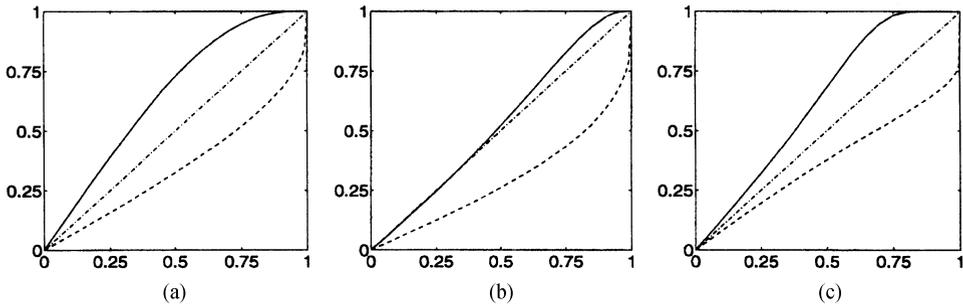


FIG. 2. (a) $d = 1$, $F = \mathcal{N}(0, 1)$, $G = \text{logistic}$; (b) $d = 1$, $F = \mathcal{N}(1, 1)$, $G = \text{logistic}$; (c) $d = 1$, $F = \mathcal{N}(0, 1)$, $G = \text{Cauchy}$.

The properties of C-C-plots listed below can be verified by straightforward calculations.

1. Suppose that the Leb-densities of F and G both belong to the same family of unimodal and spherically symmetric densities in \mathbf{R}^d with same mode and with scatter matrices $\sigma_1^2 \cdot I$ and $\sigma_2^2 \cdot I$, respectively. Then $(FG^{-1})_{\mathcal{E}}$ is convex if $\sigma_1 \geq \sigma_2$ and concave if $\sigma_1 \leq \sigma_2$ (cf. Figure 3a).

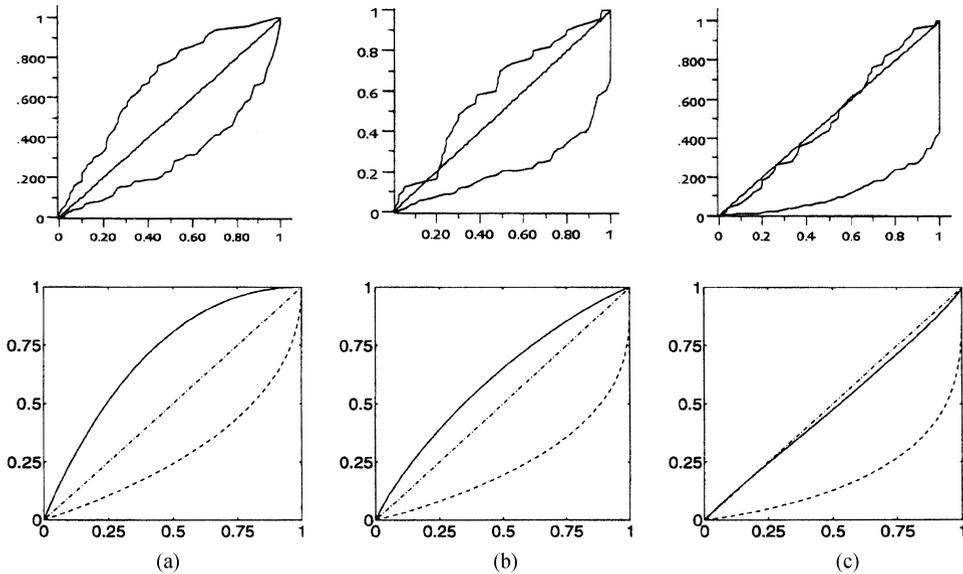


FIG. 3. Modified empirical C-C-plots based on 50 observations corresponding to the theoretical C-C-plots listed. (a) $d = 3$, $F = \mathcal{N}(0, I)$, $G = \mathcal{N}(0, 2 \cdot I)$; (b) $d = 10$, $F = \mathcal{N}(0, I)$, $G = \mathcal{N}(0, D)$, with $D = \text{diag}(1, \dots, 1, 5, 5)$; (c) $d = 10$, $F = \mathcal{N}(0, I)$, $G = \mathcal{N}(\mu, D)$, where D is as in Figure 3b and $\mu = (0, 0, 0, 0, 1, 1, 0, 0, 0, 0)$.

Convexity and concavity of $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$, respectively, of course also hold in other situations as can be seen in Figure 2a.

2. Suppose that F and G have spherically symmetric Leb-densities in \mathbf{R}^d not necessarily unimodal. In this case we have the following property:

If $G(\cdot) = F(\cdot - \mu)$ for some $\mu \in \mathbf{R}^d$ then $(FG^{-1})_{\mathcal{E}} = (GF^{-1})_{\mathcal{E}}$.

If there exists a $\sigma \in \mathbf{R}$ such that $G(x) = F(\sigma \cdot x)$ for all $x \in \mathbf{R}^d$ then $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ are inverse functions (cf. Figure 2a, c; Figure 3a).

3. Let F be a probability measure on \mathbf{R}^d . For a vector $\mu \in \mathbf{R}^d$ and a real nonsingular $d \times d$ -matrix Σ let $\{F_{\mu, \Sigma}(\cdot) = F(\Sigma^{-1} \cdot - \mu)$. Then we have the following transformation property:

$(F_{0, I} F_{\mu, \Sigma}^{-1})_{\mathcal{E}} = (F_{\mu^*, \Sigma^*} F_{0, I}^{-1})_{\mathcal{E}}$ where $\mu^* = -\Sigma^{-1/2}\mu$ and $\Sigma^* = \Sigma^{-1}$ and I denotes the identity matrix.

4. If the Leb-densities f and g of F and G , respectively, are continuous at their respective modes then we have:

f and g have different modes iff both $(FG^{-1})_{\mathcal{E}}$ and $(GF^{-1})_{\mathcal{E}}$ are smaller than α for α near zero.

Note that (4) is a generalization of the case of a shift given in (2). Here the continuity assumption can be relaxed.

6. Discussion.

Other geometric structures. The presented approach has mainly discussed the d -dimensional Euclidean space. However, the method is formulated in general terms and can be applied to many other problems with different geometric structures, too. For instance, it can be applied to directional data, or to data coming from a simplex, as is the case in some economic applications. The key is to choose an appropriate class \mathcal{E} .

Composite hypothesis. If we consider a composite hypothesis consisting of a parametric family, then of course estimating the underlying parameter is a possibility. However, unfortunately, the analogous behavior of our test statistics to classical one-dimensional test statistics carries over to this situation. Hence, for composite hypothesis our test statistics with estimated parameters will no longer be distribution free.

The two sample case. In this paper only the one-sample problem is considered. The generalization to the two-sample problem seems to be straightforward by replacing the (known) hypothetical distribution G of the previous sections by the sample distribution of the second sample. Here the computational aspect is even more critical than in the one-sample case, because one has to calculate estimates for MV-sets in both samples. Hence, the modified test statistics discussed in Section 4 become even more important.

Other continuous functionals. Instead of the sup-statistics we could also consider other continuous functionals of the underlying processes like L^2 -type statistics. This would lead to test statistics of the Cramér–von Mises type. Analogous results to the one given in the present paper for the sup-statistics can easily be obtained for L^2 -type statistics, too.

7. Proofs.

PROOF OF LEMMA 1.2. The key observation is the following geometric “extremal property” of the level sets which has already been used (implicitly) in Polonik (1998). Let

$$S(F) = \{(\nu(C), F(C)) : C \text{ measurable}\} \subset \mathbf{R} \times [0, 1],$$

then level sets of f correspond to the extremal points of the (“upper half” of the) convex hull of $S(F)$. More precisely, let \tilde{F} denote the smallest concave function from \mathbf{R} to $[0, 1]$ lying above $S(F)$, that is, the least concave majorant to $S(F)$. Let further $\partial\tilde{S}(F)$ be the intersection of the extremal points of the convex hull of $S(F)$ with the graph of \tilde{F} . [The set $\partial\tilde{S}(F)$ is the above mentioned set of extremal points of the (“upper half” of the) convex hull of $S(F)$.] Now we have: $(\nu(\Gamma_f(\lambda)), F(\Gamma_f(\lambda))) \in \partial\tilde{S}(F) \forall \lambda \geq 0$, and for every set C such that $(\nu(C), F(C)) \in \partial\tilde{S}(F)$ there exists a level $\lambda \geq 0$ such that $d_\nu(C, \Gamma_f(\lambda)) = 0$, where the λ equals the left-hand derivative of \tilde{F} in the point $l = \nu(C)$. This follows from the fact (see also the Introduction) that $\Gamma_f(\lambda) \in \operatorname{argmax}\{\nu(C) : C \in \operatorname{argmin}_{C \in \mathcal{A}}\{F(C) - \lambda\nu(C)\}\}$ by using similar arguments as used in Groeneboom (1985) for an analogous representation of the Grenander density estimator. The lesson of all this is that the set $\partial\tilde{S}(F)$ characterizes the distribution F . Of course, the same holds, if instead of $S(F)$ we consider

$$S_{\mathcal{E}}(F) = \{(\nu(C), F(C)) : C \in \mathcal{E}\} \subset \mathbf{R} \times [0, 1],$$

where we know a priori that $\mathcal{L}_f \subset \mathcal{E}$: the set $\partial\tilde{S}_{\mathcal{E}}(F)$ formed analogously to $\partial\tilde{S}(F)$ characterizes the distribution F . Now, choose $\mathcal{E} = \mathcal{L}_f \cup \mathcal{L}_g$. With this choice of \mathcal{E} it follows that if (b) holds, then $S_{\mathcal{E}}(F) = S_{\mathcal{E}}(G)$ and hence $\partial\tilde{S}_{\mathcal{E}}(F) = \partial\tilde{S}_{\mathcal{E}}(G)$ which means $F = G$. \square

PROOF OF LEMMA 2.1. (a) Suppose that P_1 holds. By the definition of MV-sets we have $G(C_{G, \mathcal{E}}(\alpha)) \geq \alpha$. Hence, the first equation in P_1 gives us $F(C_{G, \mathcal{E}}(\alpha)) \geq \alpha$. It follows, again by definition of MV-sets that $\nu(C_{G, \mathcal{E}}(\alpha)) \geq \nu(C_{F, \mathcal{E}}(\alpha))$. For symmetry reasons the converse inequality follows similarly. Hence, $\nu(C_{F, \mathcal{E}}(\alpha)) = \nu(C_{G, \mathcal{E}}(\alpha))$ and the uniqueness assumption on the MV-sets gives $d_\nu(C_{F, \mathcal{E}}(\alpha), C_{G, \mathcal{E}}(\alpha)) = 0$.

As for the proof of part (b), note first that trivially P_2 implies $F(C_{G, \mathcal{E}}(\alpha)) = F(C_{F, \mathcal{E}}(\alpha))$ and $G(C_{F, \mathcal{E}}(\alpha)) = G(C_{G, \mathcal{E}}(\alpha))$ for all $\alpha \in [0, 1]$. Hence it remains to show that P_2 implies $F(C_{F, \mathcal{E}}(\alpha)) = G(C_{G, \mathcal{E}}(\alpha))$ for all $\alpha \in [0, 1]$. We show that if there exists an $\alpha_0 \in [0, 1]$ with $F(C_{F, \mathcal{E}}(\alpha_0)) \neq G(C_{G, \mathcal{E}}(\alpha_0))$ then there exists an $\alpha \in [0, 1]$ with $d_\nu(C_{G, \mathcal{E}}(\alpha), C_{F, \mathcal{E}}(\alpha)) > 0$. Let α_0 be such

that $F(C_{F, \mathcal{E}}(\alpha_0)) = \beta_F > \beta_G = G(C_{G, \mathcal{E}}(\alpha_0)) \geq \alpha_0$. Then $C_{F, \mathcal{E}}(\alpha_0)$ is a MV-set for F at all levels $\alpha \in [\alpha_0, \beta_F]$ and $C_{G, \mathcal{E}}(\alpha_0)$ is a MV-set for G at all levels $\alpha \in [\alpha_0, \beta_G]$. However, $C_{G, \mathcal{E}}(\alpha_0)$ is no longer a MV-set for G at levels $\alpha \in (\beta_G, \beta_F]$. From the uniqueness assumption it follows that either $d_\nu(C_{F, \mathcal{E}}(\alpha_0), C_{G, \mathcal{E}}(\alpha_0)) > 0$ or $d_\nu(C_{F, \mathcal{E}}(\alpha), C_{G, \mathcal{E}}(\alpha)) > 0$ for some $\alpha \in (\beta_G, \beta_F]$. For symmetry reasons the same would follow if $\beta_F < \beta_G$. \square

PROOF OF LEMMA 2.3. Since $\mathcal{L}_g \subset \mathcal{E}$, the level sets of g are MV-sets. Hence it follows from $F \in \text{CONC}_{\mathcal{E}}(G)$ that $F(\Gamma_g(\lambda)) = G(\Gamma_g(\lambda))$ for all $\lambda > 0$. Since $\mathcal{L}_f \subset \mathcal{L}_g$ we then automatically also have $G(\Gamma_f(\lambda)) = F(\Gamma_f(\lambda))$ for all $\lambda > 0$. An application of Lemma 1.2 completes the proof. \square

PROOF OF THEOREM 3.1. Write $S_{n, \mathcal{E}} = \sup_{\alpha \in [0, 1]} |(F_n - F)(C_{G, \mathcal{E}}(\alpha)) + (F - G)(C_{G, \mathcal{E}}(\alpha))|$. The $GC(F)$ -property of the class of MV-sets in \mathcal{E} for G gives $S_{n, \mathcal{E}} = \sup_{\alpha \in [0, 1]} |(F - G)(C_{G, \mathcal{E}}(\alpha))| + o(1)$ a.s., and the assertion follows from the definition of $F \in \text{CONC}_{\mathcal{E}}(G)$. \square

PROOF OF THEOREM 3.2. (a) If $F = G$ then trivially

$$(10) \quad T_{n, \mathcal{E}} = \sup_{\alpha \in [0, 1]} |(F_n - F)(C_{G, \mathcal{E}}(\alpha))| + |(F_n - F)(C_{n, \mathcal{E}}(\alpha))|,$$

and the $GC(F)$ -property of \mathcal{E} gives the assertion.

As for the proof of part (b) note that it follows from (6) that

$$|G(C_{n, \mathcal{E}}(\alpha)) - G(C_{F, \mathcal{E}}(\alpha))| = o(1) \quad \text{a.s. as } n \rightarrow \infty,$$

and hence, by using the $GC(F)$ -property of \mathcal{E} together with (5) it follows that as $n \rightarrow \infty$,

$$\begin{aligned} T_{n, \mathcal{E}} &\geq |(F_n - G)(C_{G, \mathcal{E}}(\alpha))| + |(F_n - G)(C_{n, \mathcal{E}}(\alpha))| \\ &= |F(C_{G, \mathcal{E}}(\alpha)) - G(C_{G, \mathcal{E}}(\alpha))| \\ &\quad + |F(C_{F, \mathcal{E}}(\alpha)) - G(C_{F, \mathcal{E}}(\alpha))| + o(1) \quad \text{a.s.} \end{aligned}$$

Hence, if $F \neq_{\mathcal{E}} G$ there exists an α such that the nonstochastic term on the right-hand side is positive, and the proof of part (b) is complete.

For the proof of part (c), using the stronger assumptions, it follows similarly to the proof of part (b) that as $n \rightarrow \infty$,

$$\begin{aligned} T_{n, \mathcal{E}} &= \sup_{\alpha \in [0, 1]} [|F(C_{G, \mathcal{E}}(\alpha)) - G(C_{G, \mathcal{E}}(\alpha))| \\ &\quad + |F(C_{F, \mathcal{E}}(\alpha)) - G(C_{F, \mathcal{E}}(\alpha))|] + o(1) \quad \text{a.s.} \end{aligned}$$

and the proof is complete. \square

PROOF OF THEOREM 3.6. Starting from (10) and using (i) and (ii) it follows that

$$(11) \quad T_{n,\varepsilon} = 2 \sup_{\alpha \in [0,1]} |(F_n - F)(C_{F,\varepsilon}(\alpha))| + o_P\left(\frac{1}{\sqrt{n}}\right).$$

The assertion now follows from Theorem 3.5. \square

PROOF OF PROPOSITION 4.1. Let \hat{F}_n denote the distribution corresponding to \hat{f}_n . Let $M = \sup_x f(x)$. We have the following key inequality which holds for every $\varepsilon > 0$:

$$(12) \quad \begin{aligned} & d_F(\Gamma_f(\lambda), \hat{\Gamma}_n(\lambda)) \\ & \leq F\{x: |f(x) - \lambda| \geq \varepsilon\} \\ & \quad + \varepsilon^{-1}M \left[(\hat{F}_n - F)(\hat{\Gamma}_n(\lambda)) - (\hat{F}_n - F)(\Gamma_f(\lambda)) \right]. \end{aligned}$$

This inequality is essentially the same as inequality (3.2a) in Polonik (1995). There one just has to replace F_n by \hat{F}_n and the so-called empirical generalized λ -clusters $\Gamma_{n,\varepsilon}(\lambda)$ by $\hat{\Gamma}_n(\lambda)$.

It remains to show that the assumptions assure that for each $\varepsilon > 0$ the right-hand side in (12) converges to zero stochastically uniformly in λ . The uniform convergence to zero of the first term on the right-hand side follows from the fact that f has no flat parts. As for the stochastic term, one uses the fact that pointwise stochastic convergence of density estimators imply their L_1 -convergence, and hence one has convergence in total variation, so that the nondeterministic term converges to zero stochastically, uniformly in λ . \square

Acknowledgments. I greatly appreciate the highly constructive criticisms, remarks and comments of one referee, an Associate Editor and one of the Editors on the first version of this manuscript. They led to a substantial improvement in the presentation and the accessibility of the material. I also owe sincere thanks to Martin Moser for his interest and for his willingness to discuss some aspects of the subject. His careful reading of the manuscript not only laid bare several oversights, but it also led to the revelation of a serious error in a proof in an earlier version.

REFERENCES

- ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067. [Correction (1987) **15** 428–430.]
- BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214.
- DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.
- DUDLEY, R. M. (1974). Metric entropy of classes of sets with differentiable boundaries. *J. Approx. Theory* **10** 227–236.
- DUDLEY, R. M. (1984). A course in empirical processes. *Ecole d'Été de Probabilités de Saint Flour XII. Lecture Notes in Math.* **1097** 1–142. 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.
- FOUTZ, R. V. (1980). A test for goodness-of-fit based on an empirical probability measure. *Ann. Statist.* **8** 989–1001.
- GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neymann and Jack Kiefer* (L. Le Cam and R. Olshen, eds.) 539–555. Wadsworth, Monterey, CA.
- KHMALADZE, E. V. (1981). Martingale approach to the theory of goodness-of-fit test. *Theory Probab. Appl.* **26** 240–257.
- KHMALADZE, E. V. (1988). An innovation approach to goodness-of-fit tests in \mathbf{R}^n . *Ann. Statist.* **16** 1503–1516.
- KHMALADZE, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.
- MATHERON, G. (1975). *Random Sets and Integral Geometry*. Wiley, New York.
- POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881.
- POLONIK, W. (1997). Minimum volume sets and generalized quantile processes. *Stochastic Process. Appl.* **69** 1–24.
- POLONIK, W. (1998). The silhouette, concentration functions, and ML-estimation under order restrictions. *Ann. Statist.* **26** 1857–1877.
- ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- SAWITZKI, G., DILLER, M. and FRIEDRICH, F. (1994). *Voyager*. StatLab Heidelberg. Available at <http://statlab.uni-heidelberg.de/projects/voyager/>.
- VAN DER VAART, A. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-8705
E-MAIL: polonik@wald.ucdavis.edu