Some Extensions of Tukey's Depth Function

Jian Zhang
Institute of Systems Science, Academia Sinica, Beijing
and Institute of Statistics, University Catholique de Louvain

Abstract

As the extensions of Tukey's depth, a family of affine invariant depth functions for multivariate location and dispersion are defined for multivariate ordering. It is shown that these sample depth functions converge to their population versions uniformly on (or on any compact subset of) the parameter space. The deepest points of these depth functions are affine equivariant estimates of the corresponding multivariate location and dispersion. Some of them are proved to have asymptotic breakdown points between 1/3 and 1/2 and convergence rates of $1/\sqrt{n}$ when the underlying population follows an elliptic distribution. Their asymptotic distributions are also obtained under some regular conditions. Finally, Tukey's depth function is shown to be equivalent to an empirical likelihood in certain sense.

Key words and phrases: *Depth and outlyingness, location and dispersion estimates, breakdown point, multivariate ordering.*

AMS 1991 subject classifications: *Primary 60E99; secondary 62H99.*

Short running title: *Depth Functions*

1. Introduction

The rank of a point in a dataset is one of the key concepts in the nonparametric statistics. It is the starting point for constructing other nonparametric statistics. Since there is no really natural ordering of vectors in multi-dimensional space, how to rank a point relative to a multivariate dataset has become one of fundamental problems in multivariate statistics. In the last decades much research has been made on this topic and various ways of multivariate ordering have appeared (see Tukey, 1975; Barnett, 1976; Eddy, 1985; Reiss, 1989; Liu, 1990; Donoho and Gasko, 1992; Liu and Singh, 1993; He and Wang, 1997; Koltchinskii, 1997; Rousseeuw and Hubert, 1997, Small, 1997, and the references therein). This paper concentrates on the methods which are based on the projection pursuit technique. The projection pursuit technique represents a way to overcome the curse of dimensionality problem. The idea of the projection pursuit is to reveal the interesting structures through looking at all one-dimensional views of a dataset. The degree of the interestingness is often measured by an index. In the case of multivariate ordering, such index can be view as a depth function. For a comprehensive discussion of the projection pursuit, see Huber (1985). The first projection based depth function was put forward by Tukey in 1975 to measure how deep a point lie in the data cloud. According to this depth function, the depth of a point θ in a univariate dataset is the minimum of the number of data points on the left and on the right of θ. The depth of a point $\theta \in R^p$ with respect to a $p-$dimensional dataset $X = \{x_1, x_2, \cdots, x_n\}$ is the smallest depth in any one-dimensional projection of X. Tukey's depth is affine invariant (a depth $D(X, \theta)$ is called affine invariant if for any nonsingular $p \times p$ matrix A and any $p-$vector v, $D(AX + v, A\theta + v) = D(X, \theta)$). The contours of Tukey's depth is analogous to quantiles in the univariate case and thus can be used to construct $L-$estimates for location and dispersion. In particular, the deepest point provides an affine equivariant multidimensional median estimate. In the context of robust theory, the breakdown point is one of important quantities to measure the global robustness of a statistic. It is defined as the smallest amount of contamination that makes the statistic meaningless. For example, the asymptotic breakdown point of the univariate median estimate is 1/2. However, for $p > 1$, Donoho and Gasko (1992), and He and Wang (1997) showed that Tukey's multidimensional median estimate has 1/3 asymptotic breakdown point under elliptic symmetry.

Tukey's depth are discrete and it uses only the qualitative information contained in the multivariate ordering. In practice, one is not only interested in the ranking of the data point $x_1, \cdots, x_n$ expressed in numbers $1, 2, \cdots, n$ but also in the total distribution information contained in $x_1, \cdots, x_n$ (see Plackett, 1976). In the multidimensional case, to obtain an estimate of a higher breakdown point, some more quantitive depth functions are needed (see, Stahel, 1981; Donoho, 1982; Maronna, Stahel and Yohai, 1992; Tyler, 1994; Zhang and Li, 1995; Rousseeuw and Hubert, 1997). In this paper, motivated by the relation between Tukey's depth and the M-estimating equation and by the suggestion of Plackett (1976), we introduce a family of such depth functions by incorporating some auxiliary functions into Tukey's depth notion. It turns out Tukey's depth function is the special case of this new family of depth functions. For the location, the new depth functions can be used to rank both the data points and the values in the parameter space. However, for the dispersion, the new depth functions are used only for ranking the values in the parameter space. This is similar to the regression depth of Rousseeuw and Hubert (1997). It is shown that the new depth functions are affine invariant, consistent strongly and uniformly with respect to the compact subset of the sample or parameter space, and converge weakly to the maximum of a Gaussian process. The deepest points of these depth functions can be taken as the affine equivariant estimates of multivariate location and dispersion. Most of them are proved to have at least 1/3 asymptotic breakdown points under elliptic symmetry. Furthermore, we show that these estimates converge weakly to some functionals of Gaussian processes and have convergence rates of $1/\sqrt{n}$.

Note that depth functions are semiparametric or nonparametric in nature. Another well-known tool to rank the values in the parameter space is the likelihood contours. It is natural ask whether the depth contours can be interpreted as some semiparametric or nonparametric likelihood contours. In this paper, we show that Tukey's depth contours are equivalent to the empirical likehood contours in some sense.

The rest of paper is arranged as follows. In Section 2, to motivate the definitions of the depth functions discussed in this paper, we first discuss the relations among depth, outlyingness and M-estimating equations. After the definition, we give some properties of the new depth functions. In Section 3, we investigate the breakdown bahavior and asymptotic properties of the multivariate location and dispersion estimates based on these new depth functions. In Section 4, we propose a concept of projection based empirical likelihood. Then, we show the relation between Tukey's depth function and the projection based empirical likelihood. The proofs of the main results are presented in Section 5.

2. A family of depth functions

2.1. Depth, outlyingness and M-estimating equations

Let $X = (x_1, x_2, \cdots, x_n)$ be a $p-$dimensional dataset sampled from distribution function F, F_n be the empirical distribution function of X. For each a with $||a|| = 1$, F^a and F_n^a denote the theoretical and empirical distribution functions of the projected sample in direction a. Let P, P_n, P^a, and P_n^a denote the probability measures induced by F, F_n, F^a and F_n^a, respectively. I_A denotes the indicator function of set A. Let $D_{Tn}(\cdot)$ denote Tukey's depth function defined in the last section. Then, we have

$$
\begin{aligned}
D_{Tn}(\theta) &= n \min_{||a||=1} F_n^a(a^\tau \theta) \\
&= \min_{||a||=1} \min\{\sum_{i=1}^n I_{[a^\tau(x_i-\theta)\le 0]}, \sum_{i=1}^n I_{[a^\tau(x_i-\theta)\ge 0]}\}.
\end{aligned}
$$

From the definition, it is easy to see that Tukey's depth function describes the smallest "distance" from θ to the boundary of the dataset when θ is located in the covex hull of X. The deepest point of Tukey's depth characterizes the centre of X. If we view $I_{[a^\tau(x_i-\theta)\le 0]}$ as the absolute score to quantify the contribution of $a^\tau(x_i-\theta)$ to the depth of θ, then we have the following extensions:

$$
\begin{aligned}
D_{\psi T1n}(\theta) &= \min_{||a||=1} \min\{\sum_{i=1}^n |\psi(a^\tau(x_i-\theta)/s(a^\tau X))| I_{[a^\tau(x_i-\theta)\le 0]}, \\
&\qquad \sum_{i=1}^n |\psi(a^\tau(x_i-\theta)/s(a^\tau X))| I_{[a^\tau(x_i-\theta)\ge 0]}\}, \\
D_{\psi T2n}(\theta) &= \min_{||a||=1} \min\{\sum_{i=1}^n w_i(a^\tau(x_i-\theta)/s(a^\tau X)) I_{[a^\tau(x_i-\theta)\le 0]}, \\
&\qquad \sum_{i=1}^n w_i(a^\tau(x_i-\theta)/s(a^\tau X)) I_{[a^\tau(x_i-\theta)\ge 0]}\},
\end{aligned}
\tag{2.1}
$$

where ψ is a nondecreasing function defined on R^1 with $\psi(\infty) > 0$, $\psi(-\infty) < 0$, $s(\cdot)$ is a location invariant and scale equivariant estimate of scale, and for any sequence $\{z_i\}$ of R^1,

$$w_i(z_i) = |\psi(z_i)| / \sum_{j=1}^n |\psi(z_j)|, i = 1, \cdots, n.$$

Donoho and Gasko (1992) introduced the notion of outlyingness of θ with respect to X. Unlike the notion of depth, they use the distance between θ and the centre of X in the worst one-dimensional projection. Their outlyingness function can be expressed as

$$O_{dn}(\theta) = \max_{||a||=1} |a^\tau \theta - Med(a^\tau X)| / MAD(a^\tau X)$$

where Med denotes median and MAD denotes median absolute deviation from the median.

Comparing two definitions, we conclude that depth and outlyingness can be viewed as two "dual" concepts. Similar opinion can be found in Donoho and Gasko (1992), Liu and Singh (1993), He and Wang (1997) and Rousseeuw and Hubert

(1997). That is, we can generate an outlyingness function from a depth function only by a decreasing tranformation and vice versa. In particular, the deepest point of the depth function is corresponding to the least outlyingness point of the outlyingness function. For instance, the outlyingness function

$$n - 2D_{Tn}(\theta) = \max_{||a||=1} |\sum_{i=1}^{n}[sign(a^\tau x_i - a^\tau \theta) + I_{[a^\tau x_i - a^\tau \theta = 0]}]| \tag{2.2}$$

is a dual of D_{Tn}, where $sign(z)$ is the sign function. The depth function $lcent(\theta, X)$ in Rousseeuw and Hubert (1997) is a dual of O_{dn}. More examples can be found in Liu and Singh (1993) and He and Wang (1997). The two depth functions are called equivalent if there exists a strictly increasing transformation between them. This implies that the depth contours defined in He and Wang (1997) are invariant when the depth function is subjected to a strictly increasing transformation.

The outlyingness concept is essentially related to M-estimating equation. To see this, we consider a univariate dataset $Z = (z_1, \cdots, z_n)$ of size n from distribution $G(z - \beta_o)$, where $\beta_o \in R^1$ is unknown and G (the shape of the distribution) is not precisely known. We want to estimate β_o. Let ψ be a nondecreasing function defined on R^1 with $\psi(\infty) > 0$, $\psi(0) = 0$ and $\psi(-\infty) < 0$. Then, the M-estimate associated with ψ is defined as the solution of the equation

$$\sum_{i=1}^{n} \psi(z_i - \beta) = 0.$$

It is obvious that

$$O_{\psi n}(\beta, Z) = |\sum_{i=1}^{n} \psi(z_i - \beta)|$$

is a measure of the outlyingness of β relative to Z. And the M-estimate is just the least outlyingness point. Note that, for each i, $\psi(z_i - \beta)$ is a score to measure the contribution of the term $z_i - \beta$ to the outlyingness of β. So, such outlyingness measure is equal to the difference between the two summations of the absolute scores of the data points on the left and the right of θ. Obviously for a $p-$vector θ, its outlyingness function relative to X can be defined to be the outlyingness in the worst one-dimensional projection. That is,

$$O_{\psi n}(\theta) = \max_{||a||=1} O_{\psi n}(a^\tau \theta / s(a^\tau X), a^\tau X / s(a^\tau X)) \tag{2.3}$$

where $s(\cdot)$ is a location invariant and scale equivariant estimate of scale.

$n - 2D_{Tn}(\theta)$ in (2.2), the dual of Tukey's one-dimensional depth, is equal to $O_{sign_1 n}(\theta)$ where $sign_1(z) = -1$, when $z \le 0$; $= 1$, when $z > 0$. In this sense, both depth functions obtained from (2.1) and (2.3) can be viewed as the extensions of Tukey's depth function. However, in general the depth functions in (2.1) are different from those obtained from M-estimating equations.

Note that the M-estimate can be defined as the solution of the optimization of a certain discrepancy function called $\rho-$function. Similarly, we define the following outlyingness function

$$O_{\rho n}(\theta) = \max_{||a||=1} \sum_{i=1}^{n} \rho(a^\tau (x_i - \theta)/s(a^\tau X)). \tag{2.4}$$

By a linear transformation, the commonly used ρ function satisfies $(B\rho)$ or $(U\rho)$.

$(B\rho)$. $\rho(x)$ attains its minimum 0 at $x = 0$; ρ is nonincreasing for $x < 0$ and nondecreasing for $x > 0$; furthermore, $\rho(x) \to 1$ as $|x| \to \infty$.

$(U\rho)$. $\rho(x)$ attains its minimum 0 at $x = 0$; $\rho(x)$ is centrosymmetric and is nondecreasing for $x > 0$; $\lim_{|x|\to\infty} \rho(x) = \infty$; $\psi = \rho'$ is continuous in R^1, and there exists $x_0 \ge 0$ such that ψ is nondecreasing in $(0, x_0]$ and nonincreasing in (x_0, ∞).

2.2. Unified form and properties

Unifying the definitions in the above section, we have the following unified forms of the projection based depth and outlyingness functions. For each i, set $z_i(a, \theta) =$

$a^\tau(x_i - \theta)/s(a^\tau X)$. For each a, let g_a be a suitably chosen univariate function. g_a may depend on X and θ. Then the projection based outlyingness function of location induced by g_a is defined by

$$O_{gn}(\theta) = \max_{||a||=1} |\sum_{i=1}^{n} g_a(z_i(a,\theta))|.$$

The corresponding depth function D_{gn} is defined by

$$D_{gn}(\theta) = \frac{1}{1 + O_{gn}(\theta)/n}.$$

Special cases:

(1) $D_{\psi T1n}$ in (2.1) with ψ being bounded and odd, is equivalent to D_{gn} with $g_a(z) = \psi(\infty) - |\psi(z)|I_{[z\leq 0]}$. Let $g_a = \psi$ or ρ, we obtain the depth functions derived from (2.3) and (2.4).

(2) *A smooth version of Tukey's depth function.* As an example, we choose $g(z) = (1 - \exp(-cz))/(1 + \exp(-cz))$ with a tuning constant $c > 0$. Then, D_{gn} can be viewed as a smooth version of Tukey's depth function.

(3) *Depth functions associated with quantile functions.* For $0 < \alpha < 1$, let $g(z) = -\alpha I_{[z\leq 0]} + (1-\alpha)I_{[z>0]}$ and $g(z) = (1+\alpha)I_{[z\leq 0]} + \alpha I_{[z>0]}$, respectively, then we have two depth functions associated with $(1-\alpha)-$ quantile function. Like Tukey's depth function, these depth functions don't need a prelimiary univariate scale estimate.

(4) *Depth functions based on trimmed statistis.* Let $z_{(1)}(a,\theta) \leq \cdots \leq z_{(n)}(a,\theta)$ denote the order statistics of $z_1(a,\theta), \cdots, z_n(a,\theta)$. For $0 < \alpha < 1/2$, let

$$g_a(z) = zI_{[z_{([n\alpha])}(a,\theta)\leq z\leq z_{([n(1-\alpha)])}(a,\theta)]},$$

where $[s]$ stands for the integer part of s. Then, we obtain the depth functions based on the trimmed statistics. In particular, let $\alpha = 1/2$, we recover the location outlyingness and depth functions in Donoho and Gasko (1992), and Rousseeuw and Hubert (1997).

In the following, we focus on the case that g_a is independent of a and X. For completeness, the following propositions or theorems will include the case that g is the sign function and the corresponding results obtained by Donoho and Gasko (1992), He and Wang (1992) and Nolan (1992).

Let $s(F^a)$ denote the population version of the $s(a^\tau X)$. Set $z(a,\theta) = a^\tau(x - \theta)/s(F^a)$. Then, the population version of $D_{gn}(\theta)$ can be defined by

$$D_g(\theta) = \frac{1}{1 + |Pg(z(a,\theta))|}.$$

Let $x_{(1)}, \cdots, x_{(n)}$ be order statistics according to the corresponding depths $D_{gn}(x_{(1)}) \geq \cdots \geq D_{gn}(x_{(n)})$. Then, for $0 < \alpha < 1$, the $\alpha-$th depth contour can be constructed by

$$C_n(\alpha) = \{x : D_{gn}(x) = D_{gn}(x_{([\alpha n])})\}.$$

Proposition 2.1. (i) $D_{gn}(\theta)$ is affine invariant.

(ii) Suppose that $\max_{||a||=1} |s(a^\tau X) - s(F^a)| \to 0$, $0 < \min_{||a||=1} s(F^a) \leq \max_{||a||=1} s(F^a) < \infty$, and for any $s_1 > 0, s_2 > 0$, $P|g(s_1||x|| + s_2)| < \infty$. Suppose that g is a monotone function or satisfies one of conditions $(B\rho)$ and $(U\rho)$. Then, $D_{gn}(\theta)$ converges to $D_g(\theta)$ almost surely and uniformly on any compact subset of R^p (or on R^p when g is bounded). Futhermore, suppose that $P|g(s_1||x|| + s_2)|^2 < \infty$. Then, for each θ, $\sqrt{n}(D_{gn}(\theta) - D_g(\theta))$ converges weakly to the functional of a Gaussian process provided that

$$P\{g(a^\tau(x-\theta)/s) - g(b^\tau(x-\theta)/t)\}^2 \to 0$$

as $||a - b|| + |s - t| \to 0$.

Corollary. Suppose that $\max_{||a||=1} |s(a^\tau X - s(F^a))| \to 0$, $0 < \min_{||a||=1} s(F^a) \leq \max_{||a||=1} s(F^a) < \infty$. Assume that F has an elliptic density

$$|\Sigma_o|^{-1/2} f_o((x-\theta_o)^\tau \Sigma_o^{-1}(x-\theta_o)) > 0, x \in R^p$$

where Σ_0 is positive definte. Let F_{mo} denote the marginal distribution function of $\Sigma_o^{-1/2}(x-\theta_o)$. Let g be a bounded, odd function satisfying for any $s_1 > 0, s_2 > 0$, $P|g(s_1||x|| + s_2)| < \infty$ and $|\int g(z+v)dF_{mo}(z)|$ is strictly increasing with respect to $v \geq 0$. Then, under the Hausdorff distance, for $0 < \alpha < 1$, the $[n\alpha]-$th depth contour $C_n(\alpha)$ converges almost surely to an ellipsoid of the form $\{x \in R^p : (x-\theta_o)^\tau \Sigma_o^{-1}(x-\theta_o) = q(\alpha)\}$ where $q(\alpha)$ is a constant which depends on α, f_o and g.

Remark 2.1. $D_g(\theta)$ inherits the monotonicity property of Tukey's depth function. Suppose that g is nondecreasing odd function. If the underlying distribution is spherical, then $D_g(\theta)$ is monotonically decreasing along any fixed ray stemming from the center of the distribution. The definition of monotonicity of depth function can be found in Liu and Singh (1993).

With the same motivation of the location depth functions, we can define the projection based depth function of dispersion. To measure the scale alone, we use two methods to filter out the location effect before defining the depth function of dispersion. One is using the prelimiary location estimate, another is using a U-statistic. Croux, Rousseeuw and Hössjer (1994) used a similar U-statistic to construct the generalized $S-$Estimators of regression parameters.

Let $\mu(\cdot)$ be affine equivariant univariate location estimate. For each positive $p \times p$ matrix Σ, its outlyingness with respect to X is defined by

$$O_{\mu g n}(\Sigma) = \max_{||a||=1} |\sum_{i=1}^{n} g((a^\tau x_i - \mu(a^\tau X))/\sqrt{a^\tau \Sigma a})|.$$

The depth function of Σ is defined by

$$D_{\mu gn}(\Sigma) = \frac{1}{1 + O_{\mu gn}(\Sigma)/n}.$$

The population version is denoted by $D_{mg}(\Sigma)$. Without $\mu(\cdot)$, the outlyingness of Σ can be defined by

$$O_{ugn}(\Sigma) = \max_{||a||=1} |\sum_{i<j} g(a^\tau(x_i - x_j)/\sqrt{a^\tau \Sigma a})|.$$

The corresponding depth function of Σ is defined by

$$D_{ugn}(\Sigma) = \frac{1}{1 + O_{ugn}(\Sigma)/n}.$$

The limit version is denoted by $D_{ug}(\Sigma)$.

Two simple analogs of Tukey's depth function could be constructed in the dispersion case by letting $g(z) = sign(|z| - 1)$.

Proposition 2.2. D_{ugn} is affine invariant. Suppose g is a bounded non-decreasing function. Then, $D_{ugn}(\Sigma)$ converges to $D_{ug}(\Sigma)$ uniformly for all positive $p \times p$ matrices. Suppose there exists a location functional $\mu(F^a)$ such that $\max_{||a||=1} |\mu(a^\tau X) - \mu(F^a)|$ tends to zero almost surely. Then, $D_{\mu gn}(\Sigma)$ converges to $D_{\mu g}(\Sigma)$ almost surely and uniformly for all positive $p \times p$ matrices.

Remark 2.2. The limit versions $D_{ug}(\Sigma)$ and $D_{\mu g}(\Sigma)$ also have certain monotonicity property. Take $D_{ug}(\Sigma)$ as an example. Assume that the underlying distribution is spherical. Let P_{mo} denote the marginal distribution. Suppose g is a bounded even function and $|(P_{mo} \times P_{mo})g((z_1 - z_2)/s)|$ is increasing for $s \geq 1$ and is decreasing for $0 < s \leq 1$. Then, $D_{ug}(\Sigma_2) \leq D_{ug}(\Sigma_1)$ when $\Sigma_1 - I_p$ and $\Sigma_2 - \Sigma_1$ are positive definite; $D_{ug}(\Sigma_2) \geq D_{ug}(\Sigma_1)$ when $I_p - \Sigma_2$ and $\Sigma_2 - \Sigma_1$ are positive definite.

3. Some properties of the deepest points

Suppose $X = (x_1, \cdots, x_n)$ is an i.i.d. sample from an elliptic distribution $F = F(\cdot; \theta, \Sigma)$. The location θ and dispersion Σ are the unknown parameters of interest. There are several kinds of estimates of the multivariate location and dispersion (see Huber, 1981; Stahel, 1981; Donoho, 1982; Rousseeuw, 1985; Maronna, Stahel and Yohai, 1992; Tyler, 1994; Zhang and Li, 1995; Rousseeuw and Hubert, 1997). The deepest points of the depth functions defined in the last section give some alternative estimates for multivariate location and dispersion. They are the extensions of Tukey's halfspace median estimate.

Location estimate: $\hat{\theta} = \hat{\theta}(g, X) = argmax_\theta D_{gn}(\theta)$. Its population version:

$$argmax_\theta D_g(\theta).$$

Here θ runs over R^p.

Dispersion estimate with μ: $\hat{\Sigma}_\mu = \hat{\Sigma}(\mu, g, X) = argmax_\Sigma D_{\mu gn}(\Sigma)$. Its population version: $\Sigma_\mu(F) = argmax_\Sigma D_{\mu g}(\Sigma)$. Here Σ runs over all the $p \times p$ positive definite matrices.

Dispersion estimate without μ: $\hat{\Sigma} = \hat{\Sigma}(g, X) = argmax_\Sigma D_{ugn}(\Sigma)$. Its population version: $\Sigma(F) = argmax_\Sigma D_g(\Sigma)$. Here Σ runs over all the $p \times p$ positive definite matrices.

It is obvious that all these estimates are affine equivariant. That is, for any $p \times p$ nonsingular matrix A and p-vector v,

$$\begin{aligned}
&\hat{\theta}(g, AX + v) = A\hat{\theta}(g, X) + v;\\
&\hat{\Sigma}(\mu, g, AX + v) = A^\tau \hat{\Sigma}(\mu, g, X)A;\\
&\hat{\Sigma}(g, AX + v) = A^\tau \hat{\Sigma}(g, X)A.
\end{aligned}$$

Since these estimates are affine equivariant, throughout the remainder of the paper we assume that the true values of location and dispersion are 0 and I_p (the identity matrix), respectively. That is, F is a spherical distribution. Let F_o denote $F(\cdot, 0, I_p)$.

Under F_o, the distribution of $a^\tau x$ is independent of the direction a. Let P_{mo} denote the probability measure induced by the marginal distribution F_{mo}, and P_o denote the probability measure induced by F_o. For simplicity, the expectation of h under P_o is written as $P_o h$. "$\xrightarrow{d}$" means convergence in distribution

3.1. Asymptotics

The above estimates have decent asymptotic properties. Under some regular conditions, they are the consistent estimates of the corresponding multivariate location and dispersion. Some of them have $n^{-1/2}$ rates of convergence. These properties are presented in the following two theorems. Define

$(G0)$: For any nonnegative constants s_1 and s_2, $\int |g(s_1||x|| + s_2)| dP_o(x) < \infty$.

$(G1)$: $\int g(z-\beta) dP_{mo}(z) = 0$ if and only if $\beta = 0$.

$(G2)$: $|\int g(z-\beta) dP_{mo}(z)|$ attains the mimimum only at $\beta = 0$.

$(G3)$: $\int g((z_1 - z_2)/s) dP_{mo}(z_1) dP_{mo}(z_2) = 0$ if and only if $s = 1$.

$(G4)$: $\int g(z/s) dP_{mo}(z) = 0$ if and only if $s = 1$.

Theorem 3.1.1. (i) Suppose $\max_{||a||=1} |s(a^\tau X) - 1| \to 0$. Then, $\hat{\theta}(g, X)$ is a consistent estimate of θ when g is monotone and satisfies $(G0)$ and $(G1)$; or g satisfies $(B\rho)$ and $(G2)$; or g satisfies $(G0)$, $(U\rho)$ and $(G2)$.

(ii) If g is bounded and monotone, and satisfies $(G3)$, then $\hat{\Sigma}(g, X)$ is a consistent estimate of Σ.

(iii) If g is bounded and monotone, and satisfies $(G4)$, and $\max_{||a||=1} |\mu(a^\tau X)| \to 0$ almost surely, then $\hat{\Sigma}(\mu, g, X)$ also is a consistent estimate of Σ.

Remark 3.1.1. (1) If the density of F_{mo} is centrosymmetric and strictly decreasing in $z \geq 0$, and g satisfies $(B\rho)$ or $(C\rho)$, then $(G2)$ holds.

(2) Let $g = \psi(\infty) - |\psi(z)| I_{[z \leq 0]}$, where $\psi(z)$ is a bounded, odd and nondecreasing function. Suppose the density of F_{mo} is centreosymmetric and is strictly decreasing as $|z| \to \infty$. Then, g satisfies $(G2)$.

Theorem 3.1.2. Suppose that $\sqrt{n}(s(a^\tau X) - 1) = \sum_{i=1}^n sc(a^\tau x_i)/\sqrt{n} + o_p(1)$ and converges weakly to a normal variable $W_s(a)$. Suppose that g is a nondecreasing function satisfying: for any nonnegative constants s_1 and s_2, $P_o g(s_1||x|| + s_2)^2 < \infty$; $P_o(g(a^\tau(x-\theta)/s) - g(b^\tau(x-\theta_1)/t))^2 \to 0$ as $||a-b|| + ||\theta - \theta_1|| + |s-t| \to 0$; $|P_{mo} g(z)| = 0$; $P_o g(a^\tau(x-\theta)/s)$ has continous derivatives (with respect to (θ, s))

$$\Psi_1^a(\theta, s) = \partial P_o g(a^\tau(x-\theta)/s)/\partial s, \quad \Psi_2^a(\theta, s) = \partial P_o g(a^\tau(x-\theta)/s)/\partial s$$

with $\Psi_1^a(0,1) = d(a)a, \min_{||a||=1} |d(a)| > 0$. Then,

(i) $\hat{\theta}(g, X) = 0_p(1/\sqrt{n})$;

(ii) $\sqrt{n}\hat{\theta}(g, X)$ converges weakly to $argmin_u \max_{||a||=1} |W(g,a) + u^\tau \Psi_1^a(0,1) + W_s(a)\Psi_2^a(0,1)|$ provided that the argmin is unique with probability 1, where $W(g,a)$ is a Gaussian processes with zero means and covariance

$$EW(g,a)W(g,b) = P_o g(a^\tau x) g(b^\tau x) - P_o g(a^\tau x) P_o g(a^\tau x).$$

Let

$$\mathcal{F} = \Big\{ g\big(\tfrac{a^\tau(x_1 - x_2)}{\sqrt{a^\tau \Sigma a}}\big) : ||a|| = 1, \Sigma \text{ is any } p \times p \text{ positive matrix} \Big\}$$
$$P_o\mathcal{F} = \Big\{ \textstyle\int g\big(\tfrac{a^\tau(x_1 - x)}{\sqrt{a^\tau \Sigma a}}\big) dP_o(x_1) : ||a|| = 1, \Sigma \text{ is any } p \times p \text{ positive matrix} \Big\}$$

Put

$$U_2^n(f) = \frac{2}{n(n-1)} \sum_{i<j} f(x_i, x_j), \quad f \in \mathcal{F}.$$

Let G_p be the centered Gaussian process indexed by $L^2(R^p)$ (the space of square integrable functions) with covariance

$$EG_p(h_1)G_p(h_2) = P_o h_1 h_2 - P_o h_1 P_o h_2, \quad h_1, h_2 \in L^2(R^p).$$

Let P_o^2 denote $P_o \times P_o$.

Let l be a univariate function such that $P_{mo} l^2(z) < \infty$. Let W_μ denote a centered Gaussian process with covariance

$$EW_\mu(a_1)W_\mu(a_2) = P_o l(a_1^\tau x) l(a_2^\tau x) - P_o l(a_1^\tau x) P_o l(a_2^\tau x).$$

Theorem 3.1.3. (i) Suppose g is a bounded and nondecreasing function satisfying condition $(G3)$ and

$$P_o^2(g(a^\tau(x_1 - x_2)/\sqrt{a^\tau \Sigma_1 a}) - g(b^\tau(x_1 - x_2)/\sqrt{b^\tau \Sigma_2 b}))^2 \to 0$$

as $||a - b|| + trace(\Sigma_1 - \Sigma_2) \to 0$. $P_o^2 g(a^\tau(x_1 - x_2)/s)$ has a continous derivative $u^a(s)$ with respect to s and $\min_{||a||=1} |u^a(1)| > 0$. Then,

$$\begin{aligned}
&\hat{\Sigma}(g, X) = I_p + 0_p(1/\sqrt{n}),\\
&\sqrt{n}(\hat{\Sigma}(g, X) - I_p) \xrightarrow{d}\\
&argmin_\Delta \max_{||a||=1} |G_{P_o}(\textstyle\int g(a^\tau(x_1 - \cdot))dP_o(x_1)) + u^a(1) a^\tau \Delta a/2|
\end{aligned}$$

provided that the argmin is unique with probability 1, where Δ runs over all $p \times p$ positive definite matrices.

(ii) Suppose that g is a bounded and nondecreasing function satisfying $(G4)$ and $P_o g(a^\tau(x - t)/s)$ has continous derivatives $\Psi_1^a(\theta, s), \Psi_2^a(\theta, s)$ defined in Theorem 3.1.2 with $\Psi_1^a(0, 1) = d(a)a$, $\min_{||a||=1} |d(a)| > 0$. Assume that under P_o, $\mu(a^\tau X) = \sum_{i=1}^n l(a^\tau x_i)/n + o_p(1/\sqrt{n})$. Then,

$$\begin{aligned}
&\hat{\Sigma}(\mu, g, X) = I_p + 0_p(1/\sqrt{n}),\\
&\sqrt{n}(\hat{\Sigma}(\mu, g, X) - I_p) \xrightarrow{d}\\
&argmin_\Delta \max_{||a||=1} |W(g, a) + \Psi_1^a(0, 1)^\tau W_\mu(a) + \Psi_2^a(0, 1) a^\tau \Delta a/2|,
\end{aligned}$$

provided that the argmin is unique, where Δ runs over all $p \times p$ positive definite matrices, and $W(g, a)$ is defined in Theorem 3.1.2.

Remark 3.1.2. The assumption that the *argmin* is unique in Theorems 3.1.2–3.1.3 seems reasonable but difficult to ckeck. Kim and Pollard (1990) and He and Portnoy (1997) provided a way to check such assumption.

3.2. Breadown behavior

Let $\mathcal{H}$ be the set of all distributions on R^p. Then, the asymptotic breakdown point of location estimate $\hat{\theta}$ at the model distribution F_o is defined by

$$\epsilon(\theta(F_o)) = \inf\{\epsilon \geq 0 : \sup_{H \in \mathcal{H}} |\theta((1-\epsilon)F_o + \epsilon H)| = \infty\};$$

the asymptotic breakdown point of dispersion estimate $\hat{\Sigma}_\mu$ at the model distribution F_o is defined by

$$\begin{aligned} \epsilon(\Sigma_\mu(F_o)) &= \inf\{\epsilon \geq 0 : \inf_{H \in \mathcal{H}} \lambda_{min}(\Sigma_\mu((1-\epsilon)F_o + \epsilon H) = 0, \\ &\qquad \text{or } \sup_{H \in \mathcal{H}} \lambda_{max}(\Sigma_\mu((1-\epsilon)F_o + \epsilon H) = \infty\}, \end{aligned}$$

where λ_{min} and λ_{max} stand for the minimum and maximum eigenvalues. Similarly, define $\epsilon(\Sigma(F_o))$. It is well-known that the highest asymptotic breakdown point that the affine equivariant estimates of location and dispersion can attain is 1/2 (see Lopuhaä and Rousseeuw, 1991).

The following theorems give the lower bounds to the asymptotic breakdown points of the new multivariate location and dispersion estimates.

Theorem 3.2.1. Suppose g is a bounded, nondecreasing and odd function. Assume that for $0 \leq \epsilon < 1/2$, for any direction a and $H \in \mathcal{H}$, $0 < s((1-\epsilon)F_o^a + \epsilon H^a) < \infty$. Then,

$$\epsilon(\theta(F_o)) \geq 1/3.$$

Remark 3.2.1. Suppose that g is a bounded and nondecreasing function. Put

$$C(\epsilon) = \sup_{H\in\mathcal{H}} \min_{\theta} \max_{||a||=1} |P_o g(a^\tau(x-\theta)/s((1-\epsilon)F_o^a + \epsilon H^a))|$$

for any $0 \le \epsilon < 1/2$. Then, $\epsilon(\theta(F_o))$ satisfies

$$\epsilon(\theta(F_o)) \ge \frac{g_u - C(\epsilon(\theta(F_o)))}{2g_u + g_l - C(\epsilon(\theta(F_o)))}$$

where

$$g_u = \max\{g(\infty), -g(-\infty)\}, \quad g_l = \min\{g(\infty), -g(-\infty)\}.$$

For example, let $g(z) = -\alpha I_{[z\le 0]} + (1-\alpha)I_{[z>0]}$, then, we have

$$\epsilon(\theta(F_o)) \ge 1/3.$$

Theorem 3.2.2. Suppose g satisfies candition $(B\rho)$. For $0 \le \epsilon < 1/2$, let

$$\begin{aligned} -A(\epsilon) &= -1 + \sup\{(1-\epsilon)\textstyle\int \rho(a^\tau x/s((1-\epsilon)F_o^a + \epsilon H^a))dP_o(x) \\ &\quad + \epsilon \textstyle\int \rho(a^\tau x/s((1-\epsilon)F_o^a + \epsilon H^a))dH(x) : ||a|| = 1, H \in \mathcal{H}\}, \\ \eta &= \min\{\epsilon \ge 0 : \epsilon \ge A(\epsilon)\}. \end{aligned}$$

Then,

$$\epsilon(\theta(F_o)) \ge \frac{\eta}{1+\eta}.$$

In particular, if the scale is known, then

$$\epsilon(\theta(F)) \ge \frac{1 - \int \rho(z)dF_{mo}(z)}{2 - \int \rho(z)dF_{mo}(z)}$$

which is near 1/2 when the tunning constant is suitably chosen (see Zhang and Li, 1997).

In the following theorem, for each $0 \le \epsilon < 1/2$, we write

$$B(\epsilon) = \sup_{H\in\mathcal{H}} \max_{||a||=1} |(1-\epsilon)P_o g(a^\tau x - \mu((1-\epsilon)F_o^a + \epsilon H^a)) + \epsilon H g(a^\tau x - \mu((1-\epsilon)F_o^a + \epsilon H^a))|.$$

Theorem 3.2.3. (i) Suppose that g is increasing towards both sides, $0 < g(\infty) = g(-\infty) < \infty$. Then,

$$\epsilon(\Sigma_\mu(F_o)) \geq \min\{\epsilon \geq 0 : B(\epsilon) \geq |(1-\epsilon)\min\{|g(0)|, g(\infty)\} - \epsilon \max\{g(0), g(\infty)\}|\}.$$

In particular, if $P_{mo} g(z) = 0$ and the location is known, then

$$\epsilon(\Sigma_0(F_o)) \geq \frac{\min\{|g(0)|, g(\infty)\}}{\max\{|g(0)|, g(\infty)\} + |g(0)| + g(\infty)}.$$

(ii) Suppose that g is increasing towards both sides, $0 < g(\infty) = g(-\infty) < \infty$, and that

$$\int g(z_1 - z_2) dP_{mo}(z_1) dP_{mo}(z_2) = 0.$$

Then,

$$\epsilon(\Sigma(F_o)) \geq 1 - \sqrt{\frac{1 + c_o}{1 + 2c_o}} > 0,$$

where $c_o = \min\{|g(0)|, g(\infty)\} / \max\{|g(0)|, g(\infty)\}$.

4. Relation with empirical likelihood

If we know the likelihood, there would be some appeal in ordering the dataset in relation to the likelihood contours. When the true likelihood is unknown, we often construct a semiparametric or nonparametric likelihood, for example, empirical likelihood, by using the auxiliary information (see Owen, 1988). The depth function is semiparametric or nonparametric in nature. So, it is natural to hope the depth function has some property of semiparametric or nonparametric likelihood. In the following, we show that Tukey's depth function is equivalent to the projection based empirical likelihood.

To begin with, we construct an empirical likelihood ratio $l(a, \theta)$ of $a^\tau \theta$ (the median of the projected data $a^\tau X$) for each direction a as follows. Consider

$$\max \sum_{i=1}^{n} \log p_i$$

subject to

$$\begin{aligned} &\textstyle\sum_{i=1}^n p_i = 1, \quad p_i \geq 0, i \geq 1; \\ &\textstyle\sum_{i=1}^n p_i I_{[a^\tau(x_i-\theta)\leq 0]} = 0. \end{aligned}$$

Let $p_i(a,\theta), i = 1, \cdots, n$ be the solution. Then,

$$\begin{aligned} l(a,\theta) &= \textstyle\sum_{i=1}^n \log p_i(a,\theta) + n \log n \\ &= -nR(F_n^a(a^\tau\theta)) \end{aligned}$$

where

$$R(z) = \log 2 + z \log z + (1-z)\log(1-z), \quad 0 < z < 1.$$

The least favorable empirical likelihood among all one-dimensional views of X, $l_p(\theta) = \min_{||a||=1} l(a,\theta) = -nR(D_{Tn}(\theta)/n)$.

Proposition 4.1. Tukey's depth function is equivalent to $l_p(\theta) = -nR(D_{Tn}(\theta)/n)$.

Similarly, we can show each projection based depth function of location quantile defined in Section 2 is equivalent to a projection based empirical likelihood. In this setting,

$$R(z) = z\log(z/(1-\alpha)) + (1-z)\log((1-z)/\alpha).$$

5. Proofs of the main results

Proof of Proposition 2.1: Set $s_{min} = \min_{||a||=1} s(F^a)$, $s_{max} = \max_{||a||=1} s(F^a)$. Let Ω_1 be a compact subset of R^p bounded by $c = \max\{||\theta|| : \theta \in \Omega_1\}$. Consider the empirical process

$$\{(P_n - P)h : h \in \mathcal{F}_1\}$$

with

$$\mathcal{F}_1 = \{g(a^\tau(x-\theta)/s) : ||a|| = 1, a \in R^p, \theta \in \Omega_1, (1-\delta)s_{min} \leq s \leq (1+\delta)s_{max}\}$$

and envelope $F_1(x) = |g((||x|| + c)/s_{min})|$. It follows from the assumption that the graphs of functions in $\mathcal{F}_1$ form a polynomial class of sets [Pollard (1984), p.17]. By Theorem 24, Lemma 25 of Pollard [(1984), p.25 and p.27], we deduce that

$$\max_{\theta \in \Omega_1} |O_{gn}(\theta) - O_g(\theta)| \leq \max\{|(P_n - P)h| : h \in \mathcal{F}_1\} \to 0$$

almost surely as $n \to \infty$. And

$$\{\sqrt{n}(P_n - P)h : h \in \mathcal{F}_1\}$$

converges weakly to the centered Gaussian process W indexed by $\mathcal{F}_1$ with covariance

$$EW(h_1)W(h_2) = Ph_1h_2 - Ph_1Ph_2.$$

When g is bounded, the same result holds for $\Omega_1 = R^p$.

For each θ fixed, let

$$A_+ = \{a : ||a|| = 1, a \in R^p, Pg(a^\tau(x-\theta)/s(F^a)) = \max_{||b||=1} |Pg(b^\tau(x-\theta)/s(F^b))|\},$$

$$A_- = \{a : ||a|| = 1, a \in R^p, Pg(a^\tau(x-\theta)/s(F^a)) = -\max_{||b||=1} |Pg(b^\tau(x-\theta)/s(F^b))|\}.$$

Then, by Lemma 3.1 of Zhang and Li (1993) as well as the representation theorem of random elements (see Pollard, 1984), we have

$$\sqrt{n}(O_{gn}(\theta) - O_g(\theta)) \xrightarrow{d} \max\{\max_{a \in A_+} W(g_a), -\min_{a \in A_-} W(g_a)\}$$

where $g_a(x) = g(a^\tau(x-\theta)/s(F^a))$.

The proof is completed by using the definitions of D_{gn} and D_g.

Proof of Corollary: It is similar to the proof of Lemma 2.5 of Donoho and Gasko (1992).

Proof of Proposition 2.2: Denote $P \times P$ by P^2. Define

$$\mathcal{F}_2 = \{g(a^\tau(x_1 - x_2)/\sqrt{a^\tau \Sigma a}) : ||a|| = 1, \Sigma \text{ is positive definite }\}$$

and

$$U_2^n(f) = \frac{2}{n(n-1)} \sum_{i<j} f(x_i, x_j), f \in \mathcal{F}_2.$$

Consider the U-process

$$\{U_2^n(f) - P^2 f : f \in \mathcal{F}_2\}.$$

By the assumption, we can show that the graphs of functions in $\mathcal{F}_2$ form a polynomial class of sets [Pollard (1984), p.17]. It follows from Corollary 3.3 of Arcones and Giné (1993)

$$\sup\{|O_{ugn}(\Sigma) - O_{ugn}(\Sigma)| : \Sigma \text{ is positive definite}\} \leq \sup\{|U_2^n(f) - P^2 f| : f \in \mathcal{F}_2\} \to 0$$

almost surely which implies

$$\sup\{|D_{ugn}(\Sigma) - D_{ugn}(\Sigma)| : \Sigma \text{ is positive definite}\} \to 0$$

almost surely.

The second part is similar to the proof of Proposition 2.1 and thus omitted.

The proof is completed.

Proof of Theorem 3.1.1: It is a consequence of Propositions 2.1 and 2.2.

Proof of Remark 3.1.1: The part (1) is similar to Zhang and Li (1997) and thus omitted. To proof the part (2), we first denote the density of F_{mo} by f_{mo}, and set

$$K(t) = \int_{-\infty}^{t} |\psi(z-t)| dF_{mo}(z).$$

When $t > 0$, by the assumption of ψ, we have

$$\begin{aligned} K(t) &= \textstyle\int_{-\infty}^{t} f_{mo}(z)dz \\ &= \textstyle\int_{-t}^{\infty} \psi(z+t) f_{mo}(z)dz \\ &> \textstyle\int_{0}^{\infty} \psi(z) f_{mo}(z)dz = K(0); \end{aligned}$$

when $t < 0$, by the assumption of f_{mo}, we have

$$\begin{aligned} K(t_1) &= \int_0^\infty \psi(z) f_{mo}(z - t_1) dz \\ &< \int_0^\infty \psi(z) f_{mo}(z - t_2) dz = K(t_2) \end{aligned}$$

for $t_1 < t_2 \leq 0$. Consequently, for any $s > 0$ fixed,

$$\begin{aligned} &\max_{||a||=1} \int (\psi(\infty) - |\psi(a^\tau (x - \theta)/s)| I_{[a^\tau x - a^\tau \theta \leq 0]} dF_o(x) \\ &= \psi(\infty) - \min_{||a||=1} K(a^\tau / s) \\ &= \psi(\infty) - K(-||\theta||) \end{aligned}$$

which attains the minimum only at $\theta = 0$.

Proof of Theorem 3.1.2: To begin with, we consider the empirical process

$$\{\sqrt{n}(P_n - P)h : h \in \mathcal{F}_3\}$$

with

$$\mathcal{F}_3 = \{g(a^\tau (x - \theta)/s) : ||a|| = 1, a \in R^p, ||\theta|| \leq c, 0 < 1 - \delta \leq s \leq 1 + \delta\}.$$

The graphs of functions in $\mathcal{F}_3$ form a polynomial class of sets [Pollard (1984), p.17]. By the theorem in Pollard (1984), we deduce that the above process converges weakly to a centered Gaussian process W indexed by $\mathcal{F}_3$ with covariance

$$EW(h_1)W(h_2) = P_o h_1 h_2 - P_o h_1 P_o h_2.$$

For any $\theta_n = o_p(1)$, by the continuity the sample path of W, we derive

$$\begin{aligned} &\max_{||a||=1} \sqrt{n} |P_n g(a^\tau (x - \theta_n)/s(a^\tau X))| \\ &= \max_{||a||=1} |\sqrt{n}(P_n - P) g(a^\tau (x - \theta)/s)|_{\theta = \theta_n, s = s(a^\tau X)} \\ &+ \Psi_1^a(0,1)\sqrt{n}\theta_n(1 + o_p(1)) + \Psi_2^a(0,1)\sqrt{n}(s(a^\tau X) - 1)(1 + o_p(1))| \\ &= \max_{||a||=1} |\sqrt{n}(P_n - P_o) g(a^\tau x) + o_p(1) \\ &+ \Psi_1^a(0,1)\sqrt{n}\theta_n(1 + o_p(1)) \\ &+ \Psi_2^a(0,1)\sqrt{n}(s(a^\tau X) - 1)(1 + o_p(1))|. \end{aligned} \tag{5.1}$$

Note that it follows from Theorem 3.1.1 that $\hat{\theta} = o_p(1)$. Let $\theta_n = \hat{\theta}$ and $\theta_n = 0$ in (5.1), respectively. Then, we have

$$\begin{aligned}
&\max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) \\
&+\Psi_1^a(0,1)\sqrt{n}\hat{\theta}(1 + o_p(1)) + \Psi_2^a(0,1)\sqrt{n}(s(a^\tau X) - 1)(1 + o_p(1))| \\
&\leq \max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) \\
&+\Psi_2^a(0,1)\sqrt{n}(s(a^\tau X) - 1)(1 + o_p(1))| = 0_p(1)
\end{aligned}$$

which together with the condition on $\Psi_1^a(0,1)$ implies $\sqrt{n}\hat{\theta} = 0_p(1)$.

To prove the claim (ii), let $\theta_n(\xi) = \xi/\sqrt{n}$ and for any compact subset Θ_o of R^p, consider the process

$$\{\sqrt{n}P_n g(a^\tau(x - \theta_n(\xi))/s(a^\tau X)) : ||a|| = 1, a \in R^p, \xi \in \Theta_o\}.$$

Invoking (5.1), we conclude that this process converges weakly to

$$\{W(h) + \Psi_1^a(0,1)^\tau\xi + \Psi_2^a(0,1)W_s(a) : h(x) = g(a^\tau x), ||a|| = 1, a \in R^p, \xi \in \Theta_o\}.$$

For $\xi \in R^p$, define

$$\begin{aligned}
Z_n(\xi) &= \max_{||a||=1} \sqrt{n}|P_n g(a^\tau(x - \theta_n(\xi))/s(a^\tau X)|, \\
Z(\xi) &= \max_{||a||=1}\{|W(h) + \Psi_1(0,1)^\tau\xi + \Psi_2(0,1)W_s(a)| : \\
&\qquad h = g(a^\tau x), ||a|| = 1, a \in R^p\}.
\end{aligned}$$

Then, for any compact subset of R^p, the continous mapping theorem in Pollard (1984) gives

$$\{Z_n(\xi) : \xi \in \Theta_o\} \xrightarrow{d} \{Z(\xi) : \xi \in \Theta_o\}.$$

Hence

$$\begin{aligned}
0_p(1) &= \sqrt{n} argmin_{\theta \in R^p} \max_{||a||=1} |P_n g(a^\tau(x - \theta)/s(a^\tau X))| \\
&= argmin_{\xi \in R^p} Z_n(\xi), \\
&\min_{\xi \in R^p} Z(\xi) \leq 0_p(1).
\end{aligned}$$

By the assumption that the minimizer of $\{Z(\xi) : \xi \in R^p\}$ is unique and by applying Theorems 2.3 and 2.7 in Kim and Pollard (1990), we complete the proof of the claim (ii).

Proof of Theorem 3.1.3: (i) and (ii) can be proved along the same line as that of Theorem 3.1.2. Take (i) as an example, for any compact subset Λ of

$$\{\Sigma : \Sigma \text{ is a } p \times p \text{ positive definite matrix}\},$$

consider the U-process

$$\{\sqrt{n}(U_2^n(f) - P_o^2 f) : f \in \mathcal{F}_\Lambda\}$$

where

$$\mathcal{F}_\Lambda = \{g(a^\tau(x_1 - x_2)/\sqrt{a^\tau \Sigma a}) : ||a|| = 1, \Sigma \in \Lambda\},$$

and $U_2^n(f)$ is defined as in the proof of Proposition 2.2. Define

$$P_o\mathcal{F}_\Lambda = \{\int f(x_1, x) dP_o(x_1) : f(x_1, x) \in \mathcal{F}_\Lambda\}.$$

Then, by the assumption and the inequality

$$\int (\int f_1(x_1, x_2) dP_o(x_1) - \int f_2(x_1, x_2) dP_o(x_1))^2 dP_o(x_2) \leq P_o^2(f_1 - f_2)^2,$$

we deduce that the $\epsilon-$covering numbers of $\mathcal{F}_\Lambda$ and $P_o\mathcal{F}_\Lambda$ satisfy the conditions of Theorem 5 in Nolan and Pollard (1988) or Theorem 4.9 in Arcones and Giné (1993). Therefore,

$$\{\sqrt{n}(U_2^n(f) - P^2 f) : f \in \mathcal{F}_\Lambda\} \xrightarrow{d} \{2G_p(h) : h \in P_o\mathcal{F}_\Lambda\}.$$

For any sequence of $\Sigma_n = I_p + o_p(1)$, we have

$$\begin{aligned}
&\sqrt{n} \max_{||a||=1} |U_2^n(g(a^\tau(x_1 - x_2)/\sqrt{a^\tau \Sigma_n a}))| \\
&= \max_{||a||=1} |\sqrt{n}(U_2^n(g(a^\tau(x_1 - x_2))) - P^2 g(a^\tau(x_1 - x_2))) + o_p(1) \\
&+ u^a(1) a^\tau \sqrt{n}(\Sigma_n - I_p) a/2(1 + o_p)|.
\end{aligned}$$

The remainder is analogous to that of Theorem 3.1.2. The proof is completed.

Proof of Theorem 3.2.1: At first, we note that, for $0 \leq \epsilon < 1/2$, $||a|| = 1, a \in R^p$,

$$P_o g(a^\tau x/s((1-\epsilon)F_o^a + \epsilon H^a)) = 0$$

which implies

$$\begin{aligned}&\sup_{H\in\mathcal{H}} \min_\Theta \max_{||a||=1} |[(1-\epsilon)P_o + \epsilon H]g(a^\tau(x-\theta)/s((1-\epsilon)F_o^a + \epsilon H^a))| \\ &\leq \epsilon g(\infty).\end{aligned} \tag{5.2}$$

On the other hand, for each $\epsilon > 0$, if there exists $\{H_n\}$ such that the minimizer, θ_n of

$$\max_{||a||=1} |[(1-\epsilon)P_o + \epsilon H]g(a^\tau(x-\theta)/s((1-\epsilon)F_o^a + \epsilon H^a))|$$

tends to ∞ as $n \to \infty$, then, in the following we show it yields

$$\liminf \max_{||a||=1} |[(1-\epsilon)P_o + \epsilon H_n]g(a^\tau(x-\theta_n)/s((1-\epsilon)F_o^a + \epsilon H_n^a))| \geq (1-2\epsilon)g(\infty). \tag{5.3}$$

To this end, for $r > 0$, set

$$B_{rn} = \{x \in R^p : |\theta_n^\tau x/||\theta_n|| - ||\theta_n||| \geq r\}.$$

For any $\delta > 0$, choose $\delta_1 = \delta/(2(1-\epsilon))$. For such δ_1, choose r such that

$$g(r) \geq g(\infty) - \delta_1.$$

Then, there exists $N(r, \delta)$, when $n \geq N(r, \delta)$,

$$P_o(B_{rn}^c) \leq \delta/(4(1-\epsilon)g(\infty)).$$

Then,

$$\begin{aligned}
&\max_{||a||=1} |[(1-\epsilon)P_o + \epsilon H_n] g(a^\tau(x-\theta_n)/s((1-\epsilon)F_o^a + \epsilon H^a)| \\
&\geq |[(1-\epsilon)P_o + \epsilon H_n] g(\theta_n^\tau x/||\theta_n|| - ||\theta_n||)| \\
&\geq (1-\epsilon)g(r)P_o(B_{rn}) - (1-\epsilon)g(\infty)P_o(B_{rn}^c) - \epsilon g(\infty) \\
&\geq (1-\epsilon)(g(\infty)-\delta_1)(1-\delta/(4(1-\epsilon)g(\infty))) \\
&-(1-\epsilon)g(\infty)\delta/(4(1-\epsilon)g(\infty)) - \epsilon g(\infty) \\
&\geq (1-2\epsilon)g(\infty) - \delta.
\end{aligned}$$

(5.3) is proved. Now combining (5.2) and (5.3), we conclude that

$$\epsilon g(\infty) \geq (1-2\epsilon)g(\infty).$$

The proof is completed.

Proof of Theorem 3.2.2: The proof follows the same line as that of Theorem 3.4 of Zhang and Li (1997).

Proof of Theorem 3.2.3: The proof of (i) is similar to that of Therorem 3.2.1 and thus omitted. To prove (ii), we first note that

$$\begin{aligned}
&\sup_{H\in\mathcal{H}} \min_\Sigma \max_{||a||=1} |\textstyle\int\int g(a^\tau(x-y)/\sqrt{a^\tau\Sigma a}) \\
&((1-\epsilon)dP_o(x) + \epsilon dH(x))((1-\epsilon)dP_o(y) + \epsilon dH(y))| \\
&\leq \max_{||a||=1} |P_o^2 g(a^\tau(x-y))| \\
&+\sup_{H\in\mathcal{H}} \min_\Sigma \max_{||a||=1} |\textstyle\int\int g(a^\tau(x-y)/\sqrt{a^\tau\Sigma a})(\epsilon(1-\epsilon)dP_o(x)dH(y) \\
&+(1-\epsilon)\epsilon dP_o(y)dH(x) + \epsilon^2 dH(x)dH(y))| \\
&\leq \epsilon(2-\epsilon)\max\{g(\infty), |g(0)|\}.
\end{aligned} \tag{5.4}$$

If there exist $\{H_i\}$ such that the minimizer Σ_i of function $(*)_i$ defined by

$$\max_{||a||=1} |\int\int g(a^\tau(x-y)/\sqrt{a^\tau\Sigma a})((1-\epsilon)dP_o(x)+\epsilon dH_i(x))((1-\epsilon)dP_o(y)+\epsilon dH_i(y))|$$

is broken down, that is

$$\max\{\lambda_{max}(\Sigma_i), 1/\lambda_{min}(\Sigma_i)\} \to \infty,$$

then, anologous to the proof of Theorem 3.2.1, we can show that

$$\lim(*)_i \geq ((1-\epsilon)^2 - 2\epsilon + \epsilon^2)\min\{g(\infty), |g(0)|\}$$

which together with (5.4) completes the proof.

Acknowledgements

The author would like to thank Professor Rousseeuw and his assistants for encouragement and helpful discussions on the depth functions. The author is especially grateful to Professors I. Gijbels and L. Simar for the helpful comments which led to the improvement of presentation. Some results of the paper were presented in the Annual Meeting of the Chinese Society of Probability and Statistics in 1994. This work is partially supported by National Natural Science Foundation of China and by the contract "Project d'Actions de Recherche Concertées" no. 93/98-164.

References

[1] Arcones, M.A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Prob.* **21** 1494-1542.

[2] Barnett, V. (1976). The ordering of multivariate data (with discussions). *J. Roy. Statist. Soc. Ser. A* **139** 318-354.

[3] Croux, S., Rousseeuw, P.J. and Hössjer (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.* **89** 1271-1281.

[4] Donoho, D.L.(1982). Breakdown properties of multivariate location estimators. *Ph.D. qualifying paper,* Department of Statistics, Harvard University.

[5] Donoho, D.L. and Gasko, M. (1992). Breakdown properties of location estimates based on half-space depth and projected outlyingness. *Ann. Statist.* **20** 1803-1827.

[6] Eddy, W.F. (1985). Ordering of multivariate data. In *Computer Science and Statistics: the Interface (L. Billard, ed.)* 25-30. North-Holland, Amsterdam.

[7] Huber,P.J.(1985). Projection Pursuit (with discussions). *Ann. Statist.* **13** 435-475.

[8] Huber, P.J.(1981). Roubust Statistics. Wiley, New York.

[9] He, X. and Wang, G.(1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495-504.

[10] He, X. and Portnoy, S.(1997). Asymptotics of the deepest line. Unpublished manuscript.

[11] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191-219.

[12] Koltchinskii, V.I.(1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435-477.

[13] Liu, R.Y.(1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405-414.

[14] Liu, R.Y. and Singh, K.(1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252-260.

[15] Lopuhaä, H.P. and Rousseeuw, P.J.(1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229-248.

[16] Maronna, R.A., Stahel, W.A. and Yohai, V.(1992). Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Anal.* **42** 141-161.

[17] Nolan, D.(1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42** 157-169.

[18] Nolan, D. and Pollard, D.(1988). Functional limit theorems for U-processes. Ann. Prob. **16** 1291-1298.

[19] Owen, A.B.(1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237-249.

[20] Plackett, R.L.(1976). Discussion of Professor Barnett's paper. *J. Roy. Statist. Soc. Ser. A* **139** 344-346.

[21] Pollard, D.(1984). Convergence of Stochastic Processes. Springer, New York.

[22] Reiss, R.D.(1989). Approximate distribution of order statistics. Springer, New York.

[23] Rousseeuw, P.J.(1985). Multivariate estimation with high breakdown point. in *Mathematical Statistics and Applications*, Vol. B, eds. W. Grossmann, G.Pflug, I. Vincze, and W. Wertz, Dordrecht. Reidel, 283-297.

[24] Rousseeuw, P.J. and Hubert, M.(1997). Regression Depth. Submitted for publication.

[25] Small, C.G.(1997). Multidimensional medians arising from geodesics on graphs. *Ann. Statist.* **25** 478-494.

[26] Stahel, W.A.(1981). Robust estimation: infinitesimal optimality and covariance matrix estimators. Ph.D. thesis, ETH, Zurich.

[27] Tukey, J.W.(1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (R.D. James, ed.) **2** 523-531, Vancouver.

[28] Tyler, D.(1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22** 1024-1044.

[29] Zhang, J. and Li, G.(1993). A new approach to asymptotic distributions of maximum likelihood ratio statistics. In *Statistical Science and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference* (Matusita, K. et al. eds) 325-336, VSP, the Netherlands.

[30] Zhang, J. and Li, G.(1995). Some advances in robust estimates and tests (in Chinese). To appear in *Adv. in Math. (China).*

[31] Zhang, J. and Li, G.(1997). Breakdown properties of location M-estimators. To appear in *Ann. Statist.*

Institute of Systems Science
Academia Sinica
Beijing 100080
China.
E-mail: jzhang@iss01.iss.ac.cn