Finite Sample "Tail Behavior" of Multivariate Location Estimators

Yijun Zuo [1]

Department of Mathematics
Arizona State University
Tempe, AZ 85287-1804
USA

E-mail: yijun.zuo@asu.edu

July, 1999 (revised April, 2000)

[1]Research partially supported by a grant (FGIA) from Arizona State University.

Abstract

A finite sample performance measure of multivariate location estimators is introduced based on "tail behavior". The tail performance of various location estimators, including multivariate monotone location estimators, multivariate non-monotone location estimators such as Tukey-Donoho halfspace median and multivariate L-estimators is investigated with respect to this measure. The interconnections among our finite sample performance measure, the notion of finite sample breakdown point, and the notion of halfspace depth are explored and illuminated. It is shown that estimators with high breakdown point or high halfspace depth have extremely appealing tail performance. The tail performance of the multivariate sample median is very appealing and also robust against underlying population distributions, while the tail performance of the multivariate sample mean is very sensitive to underlying population distributions. These findings provide new insights into the notions of the halfspace depth and breakdown point and identify the important role of tail behavior as a quantitative measure of robustness in the multivariate location setting.

AMS 1991 Classification: Primary 62F12; Secondary 62G35, 62H12.

Key words and phrases: Robustness, tail behavior, data depth, breakdown point, monotonicity, halfspace median, multivariate L-estimators, location estimators.

Short Running title: Tail behavior of location estimators.

1 Introduction

Let $T_n = T_n(X_1, \ldots, X_n)$ be an estimator of an unknown parameter θ based on a random sample $X_1, \ldots, X_n$ from a random vector X in $\mathbf{R}^d$ $(d \geq 1)$ with P_θ being the corresponding probability measure. How good is T_n as an estimator of θ? Various performance measures of T_n, including *unbiasedness* and *consistency*, have been proposed and studied. Asymptotic behavior of T_n in $\mathbf{R}^1$ has also been investigated, based on the behavior of

$$-\ln P_\theta(|T_n - \theta|) > a)/n, \quad \text{as } n \to \infty$$

for *fixed* a, see Bahadur (1967, 1971), Fu (1975), and Sievers (1978).

In practice, however, the sample size n is *finite* and *fixed*. In sharp contrast with asymptotic approach to the performance of estimators, Jurečková (1981) studied, for continuous and centrally symmetric F in $\mathbf{R}^1$, the performance of location estimator T_n based on the measure

$$\frac{-\ln P_\theta(|T_n - \theta| > a)}{-\ln(1 - F(a))}, \quad \text{as } a \to \infty$$

for *fixed* n. He, Jurečková, Koenker and Portnoy (HJKP) (1990) extended this measure to the p-dimensional regression context and explored the tail behavior of various regression estimators.

In this paper, the finite sample performance measure of Jurečková (1981) in $\mathbf{R}^1$ is generalized to higher dimensions and a tail performance measure of multivariate location estimators in $\mathbf{R}^d$ $(d \geq 1)$ is introduced. Notions of monotonicity of multivariate location estimators are also introduced. The tail behavior of multivariate monotone and non-monotone location estimators is investigated based on the tail performance measure introduced.

Our finite sample approach to the performance of multivariate location estimators offers a valuable new insight into the comparison of estimator performance particularly from a robustness standpoint. In recent work on robust and nonparametric estimation and inference, two notions, the finite sample breakdown point (Donoho (1982) and Donoho and Huber (1983)) and the halfspace depth (Tukey (1975), Donoho (1982) and Donoho and Gasko (1992)), have played crucial roles. The finite sample breakdown point has emerged as a standard quantitative assessment of global robustness of estimators. The depth functions, especially the halfspace depth function, on the other hand, have become increasingly used to develop center-outward orderings of multivariate data points, and consequently to provide foundations for new robust and nonparametric methods in multivariate data analysis and inference. One of the most significant findings in this paper is that our finite sample performance measure is intimately linked with the notion of finite sample breakdown point and the notion of halfspace depth. Estimators with high breakdown point or halfspace depth always have appealing tail performance. The lower tail performance bounds of location estimators are essentially the finite sample breakdown point or the halfspace depth of the corresponding estimators (up to a scalar $1/n$). This finding offers new insights into and deepens our understanding of the notions of breakdown point and the halfspace depth and clarifies the important role of the tail behavior as a quantitative assessment of robustness of location estimators in the multivariate location setting.

The rest of this paper is organized as follows. In Section 2, a performance measure of location estimators is introduced based on their "tail behavior". A general tail performance result is obtained for a very general class of multivariate location estimators.

In Section 3, notions of monotonicity of multivariate location estimators are introduced, and the tail behavior of multivariate monotone location estimators is studied. The interconnection between the tail behavior and the finite sample breakdown point of multivariate monotone location estimators is explored and revealed.

Section 4 is devoted to the study of the tail behavior of some non-monotone location estimators, including a leading robust depth-based multivariate location estimator—the Tukey-Donoho halfspace median, and depth-based multivariate "L-estimators" (Theorems 4.1 and 4.2). The intimate relationships among the finite sample tail behavior, the "halfspace depth", and the finite sample breakdown point of these estimators are brought to light. The tail behavior of the halfspace median is very appealing and also robust against underlying population distributions (Theorem 4.1), while the tail behavior of the multivariate sample mean, a special L-estimator, is very sensitive to the underlying population distributions (Corollary 4.2).

Section 5 offers some concluding remarks.

2 A tail performance measure and a general result

Now we generalize the tail performance measure of Jurečková (1981) in $\mathbf{R}^1$ to higher dimensions. For a *fixed* n, and unit vector u, one would like the "tail" (halfspace) probability $P_\theta(u'(T_n-\theta)>a)$ to be as small as possible for a fixed $a>0$ and to approach zero as fast as possible when $a\to\infty$. On the other hand, the behavior of this tail probability depends inherently on the behavior of the tail probability $P_\theta(u'(X-\theta)>a)$ of the underlying population distribution. Thus we introduce a measure of performance, the *finite sample tail behavior*, of T_n in $\mathbf{R}^d$ based on

$$B(T_n;u,a)=\frac{-\ln P_\theta(u'(T_n-\theta)>a)}{-\ln P_\theta(u'(X-\theta)>a)},\qquad \text{as } a\to\infty,$$

for a unit vector u and a *fixed* n. The limit above (if exists) measures the relative rate of the tail probability of the sampling distribution of T_n approaching zero w.r.t. the tail probability of the underlying population distribution. Clearly, for good estimator T_n one would expect the ratio $B(T_n;u,a)$ to be high for any unit vector u as $a\to\infty$. How high can $B(T_n;u,a)$ be as $a\to\infty$? We answer this question first for any "reasonable good" location estimators, and then for various other location estimators. Specifically, we will examine the behavior of

$$\underline{B}(T_n;a)=\inf_{\|u\|=1}B(T_n;u,a)\ \text{ and }\ \overline{B}(T_n;a)=\sup_{\|u\|=1}B(T_n;u,a),\ \text{ as } a\to\infty.$$

A location estimator T_n is *affine equivariant* if $T_n(Ax^n+b)=AT_n(x^n)+b$, for any nonsingular $d\times d$ matrix A, any vector $b\in\mathbf{R}^d$, and any set of points $x^n=\{x_1,\ldots,x_n\}$ in $\mathbf{R}^d$, where $Ax^n+b=\{Ax_1+b,\ldots,Ax_n+b\}$. T_n is *translation equivariant* if the above equality holds with A being an identical matrix. Throughout this paper, we assume that T_n is translation equivariant, and X has a distribution function of the form $F(x-\theta)$. For such T_n and X, $P_\theta(u'(T_n-\theta)>a)=P_0(u'T_n>a)$ and $P_\theta(u'(X-\theta)>a)=P_0(u'X>a)$, where P_0 is the probability measure corresponding to $X-\theta$. Thus we may assume in our following discussion that $\theta=0$, and write P for P_0. (Location parameter θ can be the point of symmetry of a centrally, angularly or even halfspace symmetric distribution F. For definitions and discussions of these symmetric distributions, see Zuo and Serfling (2000c). In our following discussion, however, symmetry is not required.)

It is interesting to note that for translation (or affine) equivariant T_n, $\liminf_{a\to\infty} \underline{B}(T_n; a)$ and $\limsup_{a\to\infty} \overline{B}(T_n; a)$ are translation (or affine) *invariant.*

For a unit vector $u \in \mathbf{R}^d$, let $x_{(u,i)}$ be a point in $x^n = \{x_1, \ldots, x_n\}$ such that the rank of $u'x_{(u,i)}$ in $u'x^n = \{u'x_1, \ldots, u'x_n\}$ is i. We have the following general result.

Theorem 2.1 *Let T_n be a translation equivariant location estimator in $\mathbf{R}^d$ $(d \geq 1)$ such that for any unit vector u (1) $u'x_{(u,1)} > 0 \Rightarrow u'T_n(x^n) > 0$ and (2) $u'x_{(u,n)} \leq 0 \Rightarrow u'T_n(x^n) \leq 0$. Then*

$$1 \leq \liminf_{a\to\infty} \underline{B}(T_n; a) \leq \limsup_{a\to\infty} \overline{B}(T_n; a) \leq n.$$

Remark 2.1 (i) Roughly speaking, the theorem says that the tails of the sampling distribution of a typical location estimator approach zero (in natural log scale) at least as fast as the tails of the given underlying population distribution and can be n times as fast, and this holds uniformly with respect to u as $a \to \infty$. (ii) The lower and upper bounds are achievable by the multivariate sample mean for certain population distributions (see Section 4). (iii) The conditions (1) and (2) in the theorem hold for typical location estimators including multivariate monotone location estimators (see Section 3), multivariate medians (see Small (1990), Zuo and Serfling (2000c)), multivariate L-estimators (see Section 4). In fact, any location estimators that are located inside the convex hull formed by the sample points satisfy these conditions. A location estimator that does not satisfy these conditions might be called a *nonfit* or *inadmissible* (see Rousseeuw and Hubert (1999), and Portnoy and Mizera (1999)). (iv) Theorem 2.1 generalizes a corresponding result of Jurečková (1981) which was established for *continuous* and *centrally symmetric* distribution F in $\mathbf{R}^1$. □

Theorem 2.1 presents general lower and upper performance bounds for a wide range of location estimators. In the next two sections, sharper lower and upper performance bounds are obtained for two large classes of location estimators.

3 Tail behavior of multivariate monotone location estimators

In this section, notions of monotonicity of multivariate location estimators are introduced and their tail behavior is investigated. It turns out that the tail behavior of multivariate monotone location estimators is closely connected with their finite sample breakdown point.

3.1 Notions of monotonicity

In one dimension, it is intuitively appealing and desirable that the value of a location estimator $T_n(x_1, \ldots, x_n)$ should not be decreasing when an observation in $\{x_1, \ldots, x_n\}$ is increasing. A location estimator $T_n(\cdot)$ in $\mathbf{R}^1$ with such a property is said to be *monotone* (see HJKP (1990), Bassett (1991)). Equivalently, $T_n(\cdot)$ in $\mathbf{R}^1$ is monotone if

$$T_n(x_1, \ldots, x_n) \leq T_n(x_1^*, \ldots, x_n^*), \quad \text{whenever } x_i^* = x_i + c_i,\ c_i \geq 0, \text{ for all } i\ .$$

In $\mathbf{R}^d$ $(d \geq 1)$, as points in $x^n = \{x_1, \ldots, x_n\}$ move along the same direction u, it seems reasonable that the projection of a location estimator $T_n(x^n)$ on the direction u should at least not decrease. We thus introduce a notion of monotonicity for multivariate location estimators.

Definition 3.1 A location estimator $T_n(\cdot)$ in $\mathbf{R}^d$ $(d \geq 1)$ is said to be *monotone* if for any set of points $\{x_1, \ldots, x_n\}$ and unit vector u in $\mathbf{R}^d$

$$u'T_n(x_1, \ldots, x_n) \leq u'T_n(x_1^*, \ldots, x_n^*), \quad \text{whenever } x_i^* = x_i + c_i u,\ c_i \geq 0, \text{ for all } i.$$

A stronger version of the notion of monotonicity could be defined as follows. $T_n(\cdot)$ is said to be *strongly monotone* if for any $\{x_1, \ldots, x_n\}$ and unit vector u in $\mathbf{R}^d$,

$$u'T_n(x_1, \ldots, x_n) \leq u'T_n(x_1^*, \ldots, x_n^*), \quad \text{whenever } u'x_i \leq u'x_i^*, \text{ for all } i.$$

It is readily seen that in $\mathbf{R}^1$ the above two monotonicity notions are equivalent to that of HJKP (1990) and other notions of monotonicity which generalize the univariate one exist.

We now present some examples of (strongly) monotone location estimators in $\mathbf{R}^d$ $(d \geq 1)$.

Example 3.1 In $\mathbf{R}^1$, define (see Serfling, 1984)

$$T_n(X^n) = \sum_{i=1}^{n_{(m)}} c_{n,i} W_{n,i},$$

where $X^n = \{X_1, \ldots, X_n\}$ is a random sample from a distribution F, $c_{n,i}, 1 \leq i \leq n_{(m)}$, are arbitrary constants, $W_{n,1}, \ldots, W_{n,n_{(m)}}$ are *ordered values* of $h(X_{i_1}, \ldots, X_{i_m})$ taken over the $n_{(m)} = n(n-1)\cdots(n-m+1)$ m-tuples $(i_1, \cdots, i_m)$ of distinct elements from $\{1, \cdots, n\}$, and $h(x_1, \cdots, x_m)$ is a given "kernel" function. Then $T_n(\cdot)$ is a (strongly) monotone location estimator for suitable kernel functions. Some examples are given in the following.

(i) The sample mean, the sample α-trimmed mean, the sample quantiles, the sample median, the well-known *Hodges-Lehmann* location estimator, defined as $T_n(X^n) = \text{Med}\{(X_i + X_j)/2, i \neq j\}$, and more generally, the generalized Hodges-Lehmann location estimator, see Serfling (1984),

$$T_n(X^n) = \text{Med}\{(X_{i_1} + \cdots + X_{i_m})/m\}, \text{ for a fixed choice of } m \geq 1,$$

where $\text{Med}\{x^n\} = (x_{(\lfloor (n+1)/2 \rfloor)} + x_{(\lfloor (n+2)/2 \rfloor)})/2$, $x_{(1)} \leq \ldots \leq x_{(n)}$ are ordered value of $x_1, \ldots, x_n$.

(ii) Let kernel function $h(x) = x$, then $T_n(\cdot)$ includes the class of *L-statistics*, which are strongly monotone location estimators for any choice of nonnegative constants $c_{n,i}, 1 \leq i \leq n$.

(iii) Let $h(x_1, \cdots, x_m)$ be some linear nonnegative combinations of $x_1, \cdots, x_m$ and $c_{n,i} = 1/n_{(m)}$, for all i. Then $T_n(\cdot)$ includes the *U-statistics*, a class of strongly monotone location estimators. (For detailed discussions of classical *L*- and *U-statistics*, see Serfling (1980).) □

Example 3.2 In $\mathbf{R}^d$ $(d > 1)$, define $T_n(X^n)$ as in Example 3.1 except that $W_{n,1}, \ldots, W_{n,n_{(m)}}$ now are ordered values of $h(X_{i_1}, \ldots, X_{i_m})$ based on some multivariate ranking and ordering method, such as methods based on statistical depth functions. (For discussion on data depth and the induced ordering, see Liu (1990), Donoho and Gasko (1992), and Zuo and Serfling (2000a,b)). Some examples of multivariate monotone location estimators are given in the following.

(i) The sample mean in $\mathbf{R}^d$ $(d > 1)$ is a monotone location estimator.

(ii) Let $h(x) = x$ and $W_{n,1}, \ldots, W_{n,n_{(m)}}$ be $X_1, \ldots, X_n$ respectively, then $T_n(\cdot)$ is a monotone location estimator for any choice of nonnegative data-independent constants $c_{n,i}$, $1 \leq i \leq n$. □

All above location estimators are strongly monotone (thus monotone). The following estimators are monotone but not strongly monotone.

Example 3.3 Let T_n be coordinatewise median or coordinatewise Hodges-Lehmann estimators in $\mathbf{R}^d$ $(d \geq 1)$. Then it is not difficult to show that T_n is monotone but not strongly monotone. There are other (coordinatewise) location estimators which are monotone but not strongly monotone. □

3.2 Tail behavior and breakdown point of monotone location estimators

In the following, general tail performance bounds of multivariate monotone location estimators are obtained. The interconnections between the tail behavior and breakdown point of multivariate monotone location estimators are brought to light. The finite sample *replacement breakdown point* (RBP) of an estimator $T_n(\cdot)$ at X^n in $\mathbf{R}^d$ is defined as (see Donoho (1982))

$$\mathrm{RBP}(T_n, X^n) = \min\{\frac{m}{n} : \sup_{X^n_m} \|T_n(X^n_m) - T_n(X^n)\| = \infty\},$$

where X^n_m denotes the contaminated dataset resulting from replacing m points of X^n with arbitrary points. In other words, $\mathrm{RBP}(T_n, X^n)$ is the minimum fraction of contaminated data which could drive the estimator T_n beyond any bounds. In the following, we assume that the RBP of T_n is "sample-independent". In fact, it can be shown that the RBP of multivariate translation equivariant monotone location estimators are sample-independent. For T_n with RBP$= m/n$, define conditions
C-Plane: if there are $n - m + 1$ x_i's in a hyperplane H, then $T_n(x_1, \ldots, x_n) \in H$,
C-Spot: if there are $n - m + 1$ x_i's coinciding with b, then $T_n(x_1, \ldots, x_n) = b$.

Theorem 3.1 *Let T_n be a translation equivariant location estimator with RBP$= m/n$. Then*

$$m \le \liminf_{a\to\infty} \underline{B}(T_n; a) \le \limsup_{a\to\infty} \overline{B}(T_n; a) \le n - m + 1,$$

if any one of the following conditions holds
(A) T_n *is monotone and satisfies* C-Plane,
(B) T_n *is strongly monotone and satisfies* C-Spot,
(C) T_n *is strongly monotone and* $\lim_{a\to\infty} \ln P(u'X > a + c)/\ln P(u'X > a) = 1$ *for any fixed c.*

Remark 3.1 (i) C-Plane is satisfied by affine equivariant T_n with RBP m/n. (ii) C-Spot is satisfied by translation and *scale* equivariant T_n with RBP m/n. (T_n is scale equivariant if $T_n(cx_1, \ldots, cx_n) = cT_n(x_1, \ldots, x_n)$ for any scalar c.) (iii) The limit in condition (C) above holds for a wide range of distributions (including multivariate normal distributions). (iv) The theorem reveals inherent connections between the finite sample tail behavior and the finite sample breakdown point of multivariate monotone location estimators. High RBP of a monotone location estimator T_n corresponds to attractive tail performance of T_n. (v) The result also provides probabilistic interpretations of the finite sample RBP based on the finite sample tail behavior of monotone location estimators. (vi) Part (C) of the theorem improves and extends a corresponding result of HJKP (1990). □

Monotonicity seems to be a reasonable assumption on location estimators and is satisfied by estimators like those in Examples 3.1, 3.2 and 3.3. A high breakdown point corresponds to a high lower tail performance bound of multivariate monotone location estimators. However, there are few known monotone location estimators which have very high breakdown point. Coordinatewise median is a monotone location estimator with the highest possible breakdown point. But here the desirable affine (even rigid-body) equivariance is sacrificed. Monotonicity is more restrictive than it appears. In fact, some leading multivariate location estimators such as multivariate medians are not monotone. In the next section, we investigate the tail behavior of some multivariate non-monotone location estimators.

4 Tail behavior of multivariate non-monotone location estimators

In this section, we explore the tail behavior of some selected leading robust depth-based multivariate location estimators. In general, these estimators are not monotone.

4.1 Tail behavior of the Tukey-Donoho halfspace median

Tukey (1975) introduced the notion of the "halfspace depth" and suggested its role in defining multivariate analogues of univariate rank and order statistics via depth-induced "contours". Further development in the halfspace depth was stimulated by Donoho (1982) and Donoho and Gasko (1992), who studied its properties and focused on the robustness properties of the associated notion of "halfspace median". Recent studies of depth include Liu (1990), He and Wang (1997), He and Portnoy (1998), Mizera (1998), Rousseeuw and Hubert (1999), Liu, Parelius and Singh (1999), Nolan (1999), Massé (1999), and Zuo and Serfling (2000a,b,d).

The *halfspace depth* of a point x in $\mathbf{R}^d$ with respect to a random sample $X^n = \{X_1, \ldots, X_n\}$ from F, denoted by $HD_n(x; X^n)$, is defined as

$$
\begin{aligned}
HD_n(x; X^n) &= \inf\{F_n(H) : H \text{ a closed halfspace, } x \in H\},\ x \in \mathbf{R}^d \\
&= \inf_{\|u\|=1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}(u'X_i \geq u'x),
\end{aligned}
$$

where F_n is the empirical distribution with mass $1/n$ on each X_i $(1 \leq i \leq n)$. Thus the halfspace depth of a point x is the minimum fraction of data points contained in a closed halfspace which contains the point x. The *halfspace median* HM_n of X^n is the point with maximum halfspace depth,

$$
HM_n(X^n) = \arg \sup_{x \in \mathbf{R}^d} HD_n(x; X^n).
$$

The halfspace depth of HM_n is bounded between $1/(d+1)$ and $1/2$, and can approach $1/2$ for continuous centrally symmetric F (see Donoho and Gasko (1992)). It is readily seen that HM_n is affine equivariant. It is, however, not monotone in $\mathbf{R}^d$ $(d > 1)$. By Theorem 2.1 the tail performance measure of HM_n is bounded between 1 and n as a approaches infinity. In the following, sharper lower and upper performance bounds are obtained.

Theorem 4.1 *Let HM_n be the halfspace median and $HD_n(HM_n(X^n); X^n) \geq D/n$ for any X^n. Then*

$$
D \leq \liminf_{a \to \infty} \underline{B}(HM_n; a) \leq \limsup_{a \to \infty} \overline{B}(HM_n; a) \leq n - D + 1.
$$

Remark 4.1 (i) The theorem holds with no specific assumptions on population distribution F. (ii) The theorem reveals intimate connections between the finite sample tail behavior and the halfspace depth of HM_n. Roughly speaking, the tails of the sampling distribution of HM_n approach zero at least D times as fast as the tails of the underlying population distribution and can be $n - D + 1$ times as fast. (iii) Since the halfspace depth is maximized at HM_n, HM_n minimizes the maximum range of the tail performance bounds by maximizing and minimizing the least and the most favorable tail performance bounds respectively. It is thus a min-max procedure. (iv) The theorem gives an interpretation of the halfspace depth based on the tail behavior of HM_n. (v) It can be shown that the RBP of HM_n is $\geq D/n$. So the tail behavior of HM_n is closely linked with its RBP as well as its halfspace depth. □

In R^1, HM_n reduces to the univariate median (Med_n). By Theorem 4.1 we have

Corollary 4.1 *Let Med_n be the univariate sample median. Then*

$$\lfloor (n+1)/2 \rfloor \le \liminf_{a\to\infty} \underline{B}(Med_n; a) \le \limsup_{a\to\infty} \overline{B}(Med_n; a) \le \lfloor (n+2)/2 \rfloor.$$

Remark 4.2 (i) Corollary 4.1 can also be derived from Theorems 3.1. Under central symmetry and other assumptions on F, HJKP (1990) and Jurečková (1981) obtained the same result. (ii) Since the RBP of Med_n is $\lfloor (n+1)/2 \rfloor / n$, thus we see again how the tail performance, the breakdown point, and the halfspace depth of Med_n are intimately connected. □

The halfspace median is robust (high RBP and tail performance bounds). Its computational burden, however, is substantial. In the following, tail behavior of a class of computationally less expensive robust location estimators is studied.

4.2 Tail behavior of multivariate L-estimators

Univariate *L-estimators*, as linear combinations of one-dimensional ordered statistics, have long been recognized as robust location estimators in one dimension with desirable properties. Statistical depth functions can provide a center-outward ordering of multivariate observations and consequently allow one to extend the concept of classical univariate L-estimators to the multivariate setting. Let $X_1, \ldots, X_n$ be a random sample from F in $\mathbf{R}^d$, and $X_{(1)}, \ldots, X_{(n)}$ be the corresponding ordered multivariate statistics based on some multivariate ranking and order methods (such as depth function based methods). T_n is said to be a multivariate L-*estimator* if it is of the form $T_n = \sum_{i=1}^n c_i X_{(i)}$, where $c_i \ge 0$, $1 \le i \le n$, and $\sum_{i=1}^n c_i = 1$. The order involved in the above order statistics and c_i's are assumed to be translation invariant. Multivariate L-estimators are thus translation equivariant. However, in general they are not monotone in R^d ($d > 1$) since c_i's are typically data dependent. By Theorem 2.1 it is seen that the tail performance measure of multivariate L-estimators is bounded below and above by 1 and n respectively as $a \to \infty$. In the following, sharper performance bounds of multivariate L-estimators are obtained for certain population distributions. It is also shown that both the lower and the upper bounds 1 and n can be attained by the multivariate sample mean.

Define $c = 1/\max_i\{c_i\}$, $c^* = \min_i\{c_i : c_i > 0\}$ and $s = \sum_i \mathbf{I}(c_i > 0)$, then $1 \le c \le s \le 1/c^* \le n$. For simplicity we assume in the following that the halfspace depth is utilized in the center-outward ordering of multivariate sample points in the multivariate L-estimators.

Theorem 4.2 *Let $T_n = \sum_{i=1}^n c_i X_{(i)}$ be a multivariate L-estimator with halfspace depth $\ge m/n$ for any X^n and c and c^* are constants.*

(i) *If for some $b, C > 0$ and $r \ge 1$,*

$$(1) \quad \lim_{a\to\infty} \frac{-\ln P(u'X > a)}{ba^r} = 1, \quad \text{and} \quad (2) \quad \int_{|u'x|>C} (P(u'X > |u'x|))^{-\alpha}\, dF(x) < \infty, \ \forall\, \alpha < 1,$$

uniformly w.r.t. unit vector $u \in \mathbf{R}^d$, then

$$\max\{m, c\} \le \liminf_{a\to\infty} \underline{B}(T_n; a) \le \limsup_{a\to\infty} \overline{B}(T_n; a) \le n - m + 1.$$

(ii) *If*

$$\lim_{a\to\infty} \frac{-\ln P(u'X > a)}{\gamma \ln a} = 1, \quad \textit{for some } \gamma > 0,$$

uniformly w.r.t. unit vector $u \in \mathbf{R}^d$, *then*

$$m \le \liminf_{a\to\infty} \overline{B}(T_n; a) \le \limsup_{a\to\infty} \overline{B}(T_n; a) \le \min\{n - m + 1, n - s + 1\}.$$

Remark 4.3 (a) It can be shown that the RBP of T_n in the theorem is $\ge m/n$, thus we again see the interconnections among the tail behavior, the halfspace depth and the breakdown point for multivariate L-estimators. (b) Distributions satisfying the conditions in (i) and the condition in (ii) of Theorem 4.2 are respectively called *light tailed* and *heavy tailed*. Light tailed distributions include multivariate normal distributions, while heavy tailed distributions include multivariate t distributions. In $\mathbf{R}^1$ condition (2) of (i) can be dropped since the condition (1) of (i) implies it. □

By Theorem 4.2, the multivariate sample mean can have the worst and the best tail performance.

Corollary 4.2 *Let* $T_n = \sum_i^n X_i/n$ *be the multivariate sample mean. Then for light* (*or heavy*) *tailed distribution* F, $\lim_{a\to\infty} \underline{B}(T_n; a) = \lim_{a\to\infty} \overline{B}(T_n; a) = n$ (or 1) *respectively.*

Remark 4.4 (i) Corollary 4.2 generalizes a univariate result (Theorem 2.2) of Jurečková (1981). (ii) It is interesting to note that the tail performance of the sample mean is sensitive to population distribution F , while the behavior of the halfspace median is robust against F (Theorem 4.1). □

5 Concluding Remarks

In the univariate setting, the tail behavior of location estimators was studied by many authors, including Huber (1968), Jurečková (1981a,b, 1885), Antoch (1984a,b), and Kušnius and Mizera (1999). (Some of these studies were before the study of breakdown point.) He, Jurečková, Koenker and Portnoy (HJKP) (1990) explored the tail behavior of various regression estimators. The underlying distribution F in these studies was usually assumed to be continuous, symmetric and (or) satisfy other conditions. In the discussion of this paper, symmetry and continuity are not required and there is usually no (or only mild) conditions on F. Many results obtained in the paper improve and extend previous univariate ones in the literature.

Lopuhaä and Rousseeuw (LR) (1991) (Theorem 4.1) made a first attempt to generalize a result of HJKP (1990) to the multivariate location setting. The tail performance measure defined in LR (1991) is based on Euclidean norm, whereas the measure in this paper is based on half spaces. Besides, T_n in LR (1991) needs to satisfy a monotone condition and F is required to be spherically symmetric. The results in this paper are more general and do not follow from LR (1991).

The connection between the tail behavior and the breakdown point (robustness) was revealed in Jurečková (1981a), HJKP (1990), and LR (1992) for location or regression estimators for continuous and symmetric F satisfying some tail conditions. This connection is extended in this paper to the general multivariate location setting. Further, a connection among the halfspace depth, the tail behavior and the breakdown point is revealed in the paper, offering a new insight into the notion of halfspace depth and clarifying the role of tail behavior as a universal measure of robustness.

Finally we comment that for a selected set of location estimators, we have studied their tail behavior. Our tail performance measure is based on half spaces, hence it works very well for (and perhaps only for) monotone and halfspace based location estimators. The tail behavior of various other estimators including multivariate M-estimators, S-estimators, and estimators of scale and scatter matrices merits investigation, and different tail performance measures seems necessary for these estimators.

Appendix: Selected Proofs

Proof of Theorem 3.1. Case (A). We show that T_n satisfies

$$u'x_{(u,m)} \leq u'T_n(x_1, \ldots, x_n) \leq u'x_{(u,n-m+1)}.$$

For a given unit vector u, let R_i^u be the rank of $u'x_i$ among $u'x_1, \ldots, u'x_n$. Let $x_i^* = x_i - x_{(u,m)}$. By the translation equivariance and monotonicity of T_n, we have

$$\begin{aligned} u'T_n(x_1, \ldots, x_n) &= u'x_{(u,m)} + u'T_n(x_1^*, \ldots, x_n^*) \\ &\geq u'x_{(u,m)} + u'T_n\left(x_1^{**} \ldots, x_n^{**}\right), \end{aligned}$$

where $x_i^{**} = x_i^*\mathbf{I}(R_i^u \leq m) + (x_i^* - c_i u)\mathbf{I}(R_i^u > m)$, $c_i = u'x_i^*$, for all i, and $\mathbf{I}(\cdot)$ is the indicator function. Since there are at least $n - m + 1$ points in $\{x_1^{**}, \ldots, x_n^{**}\}$ contained in the hyperplane $H = \{x \in \mathbf{R}^d : u'x = 0\}$, thus $u'T_n\left(x_1^{**} \ldots, x_n^{**}\right) = 0$, hence $u'T_n(x_1, \ldots, x_n) \geq u'x_{(u,m)}$. Likewise, one could show that $u'T_n(x_1, \ldots, x_n) \leq u'x_{(u,n-m+1)}$. Let $m^* = n - m + 1$. Then for any unit vector u and positive number a,

$$\begin{aligned} P(u'T_n > a) &\leq P(u'X_{(u,n-m+1)} > a) \\ &= \sum_{s=m}^{n} \binom{n}{s} \left(P(u'X > a)\right)^s \left(P(u'X \leq a)\right)^{n-s} \\ &= \left(P(u'X > a)\right)^m \sum_{s=m}^{n} \binom{n}{s} \left(P(u'X > a)\right)^{s-m} \left(P(u'X \leq a)\right)^{n-s}. \end{aligned}$$

Similarly, we have

$$P(u'T_n > a) \geq \left(P(u'X > a)\right)^{m^*} \sum_{s=m^*}^{n} \binom{n}{s} \left(P(u'X > a)\right)^{s-m^*} \left(P(u'X \leq a)\right)^{n-s}.$$

The desired result now follows.

Case (B). The desired result follows if we could show that T_n satisfies

$$u'x_{(u,m)} \leq u'T_n(x_1, \ldots, x_n) \leq u'x_{(u,n-m+1)}.$$

In the following, we only prove $u'T_n(x_1, \ldots, x_n) \geq u'x_{(u,m)}$. The proof for the other part is similar. By the translation equivariance and strongly monotonicity of T_n, we have

$$\begin{aligned} u'T_n(x_1, \ldots, x_n) &= u'x_{(u,m)} + u'T_n(x_1^*, \ldots, x_n^*) \\ &\geq u'x_{(u,m)} + u'T_n\left(x_1^{**}, \ldots, x_n^{**}\right), \end{aligned}$$

where $x_i^* = x_i - x_{(u,m)}$, $x_i^{**} = x_i^*\mathbf{I}(R_i^u \leq m)$ for all i. Since there are at least $n - m + 1$ 0's in $\{x_1^{**}, \ldots, x_n^{**}\}$, thus $T_n\left(x_1^{**}, \ldots, x_n^{**}\right) = 0$. It follows that $u'T_n(x_1, \ldots, x_n) \geq u'x_{(u,m)}$.

Case (C). Following the proof of Case (B), we have

$$u'T_n(x_1, \ldots, x_n) \geq u'x_{(u,m)} + u'T_n\left(x_1^{**}, \ldots, x_n^{**}\right),$$

with at least $n - m + 1$ 0's in $\{x_1^{**}, \ldots, x_n^{**}\}$. Since the RBP of T_n is m/n, thus for a contaminated data set $\{Y_1, \ldots, Y_n\}$ resulting from replacing $(m-1)$ 0 vectors in $\{0, \ldots, 0\}$ by $(m-1)$ arbitrary vectors, there is a positive number M such that $\|T_n(Y_1, \ldots, Y_n)\| < M$. Hence we have $u'T_n(x_1, \ldots, x_n) \geq u'x_{(u,m)} - M$. Likewise, we can show that $u'T_n(X^n) \leq u'X_{(u,n-m+1)} + M$. Now

$$\begin{aligned} P(u'T_n > a) &\leq P(u'X_{(u,n-m+1)} > a - M) \\ &= \sum_{s=m}^{n} \binom{n}{s} \left(P(u'X > a - M)\right)^s \left(P(u'X \leq a - M)\right)^{n-s} \\ &= \left(P(u'X > a - M)\right)^m \sum_{s=m}^{n} \binom{n}{s} \left(P(u'X > a - M)\right)^{s-m} \left(P(u'X \leq a - M)\right)^{n-s}. \end{aligned}$$

Similarly, let $m^* = n - m + 1$, we have

$$P(u'T_n > a) \geq \left(P(u'X > a + M)\right)^{m^*} \sum_{s=m^*}^{n} \binom{n}{s} \left(P(u'X > a + M)\right)^{s-m^*} \left(P(u'X \leq a + M)\right)^{n-s}.$$

The desired result now follows. □

Proof of Theorem 4.1. The middle inequality is trivial. By the definition of the halfspace depth, if $u'HM_n > a$ for a given unit vector u and a large positive number a, then there are at least some D sample points, say $X_{i_1}, \ldots, X_{i_D}$ in $\{X_1, \ldots, X_n\}$, such that $u'X_{i_j} > a$ for $1 \leq i_j \leq n$, $1 \leq j \leq D$, or otherwise the halfspace depth of HM_n will be less than D/n. Let $A(D, a, u, X^n) =$"there are at least some D sample points in X^n such that the projection to u of each of these points is greater than a". Thus

$$\begin{aligned} P(u'HM_n > a) &\leq P(A(D, a, u, X^n)) \\ &= \sum_{s=D}^{n} \binom{n}{s} (P(u'X > a))^s (P(u'X \leq a))^{n-s} \\ &= (P(u'X > a))^D \left(\sum_{s=D}^{n} \binom{n}{s} (P(u'X > a))^{s-D} (P(u'X \leq a))^{n-s} \right). \end{aligned}$$

The desired leftmost inequality follows. To prove the rightmost inequality, we note that if there are only some $(D-1)$ points in X^n whose projection onto a given unit vector is no greater than a given positive number a, then $u'HM_n$ must be greater than a, or otherwise the sample halfspace depth of HM_n will be less than D/n. Thus we have

$$\begin{aligned} P(u'HM_n > a) &\geq P(u'X_{i_1} \leq a, \ldots, u'X_{i_{D-1}} \leq a, u'X_{i_D} > a, \ldots, u'X_{i_n} > a, \quad 1 \leq i_j, j \leq n) \\ &= (P(u'X \leq a))^{D-1} (P(u'X > a))^{n-D+1}, \end{aligned}$$

which implies the desired rightmost inequality. $\square$

Proof of Theorem 4.2. Following the proof of Theorem 4.1, one could show that for any random sample X^n

$$m \le \liminf_{a\to\infty} \underline{B}(T_n; a) \le \limsup_{a\to\infty} \overline{B}(T_n; a) \le n - m + 1.$$

For any $\epsilon \in (0,1)$ and unit vector $u \in \mathbf{R}^d$, it follows from Markov's inequality that

$$P(u'T_n > a) \quad \le \quad \frac{E[\exp((1-\epsilon)cb(u'T_n)^r)]}{\exp((1-\epsilon)cba^r)},$$

Thus $-\ln P(u'T_n > a) \ge -\ln E[\exp((1-\epsilon)cb(u'T_n)^r)] + (1-\epsilon)cba^r$.

If we now could show that $E[\exp((1-\epsilon)cb(u'T_n)^r)]$ is bounded uniformly with respect to u for fixed n, then combining the above findings and condition (1) of part (i), we obtain part (i) of the desired result. By Hölder's inequality, we get

$$\begin{aligned} E[\exp((1-\epsilon)cb(u'T_n)^r)] &\le E[\exp((1-\epsilon)cb(\sum_i c_i|u'X_{(i)}|^r))] \\ &\le E[\exp((1-\epsilon)b(\sum_i |u'X_i|^r))] \\ &= (E[\exp((1-\epsilon)b(|u'X|^r))])^n. \end{aligned}$$

Now we show that $E[\exp((1-\epsilon)b(|u'X|^r))]$ is bounded uniformly with respect to u. By condition (1), there exists a $\beta > 0$ and $M > 0$ (independent of u) such that $\beta < 1/(1-\epsilon) - 1$ and $ba^r < -(1+\beta)\ln P(u'X > a)$ when $a > M$. We thus have

$$\begin{aligned} E[\exp((1-\epsilon)b(|u'X|^r))] &\le \exp((1-\epsilon)bM^r) + \int_{|u'x|>M} \exp((1-\epsilon)b|u'x|^r)dF(x) \\ &\le \exp((1-\epsilon)bM^r) + \int_{|u'x|>M} (P(u'X > |u'x|))^{-(1-\epsilon)(1+\beta)}dF(x), \end{aligned}$$

which is bounded uniformly with respect to u. Part (i) follows.

Now we prove part (ii) of the desired result. Let $0 < c_{(1)} \le c_{(2)} \le \ldots \le c_{(s)}$ be the ordered values of nonzero $c_1, \ldots, c_n$ and $c' = \sum_{i=2}^s c_{(i)}$. For any unit vector u and $a > 0$, we have

$$\begin{aligned} P(u'T_n > a) &\ge P\left(u'X_{(u,1)} > -a/c',\ \ldots,\ u'X_{(u,s-1)} > -a/c',\ u'X_{(u,s)} > 2a/c^*\right) \\ &= \binom{n}{s-1}(P(u'X > -a/c'))^{s-1}\left(P\left(u'X > 2a/c^*\right)\right)^{n-s+1}. \end{aligned}$$

Since

$$\frac{-(s-1)\ln P(u'X > -a/c') - \ln\binom{n}{s-1}}{-\ln P(u'X > a)} \quad \to \quad 0, \quad \text{as } a \to \infty,$$

and

$$\frac{-\ln P\left(u'X > 2a/c^*\right)}{-\ln P(u'X > a)} \quad \to \quad 1, \quad \text{as } a \to \infty,$$

uniformly with respect to u. Part (ii) follows. $\square$

Acknowledgments

The author greatly appreciates the insightful and constructive remarks of an Associate Editor and two referees, which led to distinctive improvements in the paper. He is grateful to Peter Rousseeuw for his useful comments on the notion of monotonicity, to Xuming He for his various useful suggestions and comments, to Robert Serfling for his stimulating discussions and constant encouragement, and to Dennis Young for his helpful comments. The work is partially supported by a grant (FGIA) from Arizona State University.

References

[1] Antoch, J. (1984a). Behavior of L-estimators of location from the point of view of large deviations. In *Robustness of Statistical Methods and Nonparametric Statistics* (D. Rasch and M. L. Tiku, eds) 4-8. Deutscher Verlag der Wissenschaften, Berlin.

[2] Antoch, J. (1984b). Behavior of estimators of location in non-regular cases: A Monte Carlo Study. *Proc. 3rd Prague Symp. on Asympt. Statis.* (P. Mandl and M. Hušková, eds) 185-196. Elsevier, Amsterdam.

[3] Bahadur, R. R. (1967). Rates of Convergence of Estimators and Test Statistics. *Ann. Math. Statist.* **38** 303-324.

[4] Bahadur, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM. Philadelphia.

[5] Bassett, G. W. (1991). Equivariant, monotonic, 50% breakdown estimators. *Amer. Statist.* **45** 135–137.

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

[7] Donoho, D. L., and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift foe Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157-184. Wadsworth, Belmont, CA.

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

[9] Fu, J. C. (1975). The rate of convergence of consistent point estimators. *Ann. Statist.* **3** 234-240.

[10] He, X., Jurečková, J., Koenker,R. and Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58** 1195–1214.

[11] He, X., and Portnoy, S. (1998). Asymptotics of the deepest line. In *Applied Statistical Science III: Nonparametric Statistics and Related Topics* 71-81. Nova Science Publication, New York.

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

[13] Huber, P. J. (1968). Robust confidence limits. *Z. Wahrscheinlichkeitstheorie werw. Geb.* **10** 269-278.

[14] Jurečková, J. (1981a). Tail behavior of location estimators. *Ann. Statist.* **9** 578-585.

[15] Jurečková, J. (1981b). Tail behavior of location estimators in non-regular cases. *Comment. Math. Univ. Carolinae* **22** 365-375.

[16] Jurečková, J. (1985). Tail behavior of L-estimators and M-estimators. *Proc. 4th Pannonian Symposium* (W. Wertz, ed) Vol. 1, 205-217.

[17] Kušnier, J. and Mizera, I. (1999). Tail behavior and breakdown properties of equivariant estimators of location. To appear in *Ann. Inst. Statist. Math.*

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

[19] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate Analysis by data depth: descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **27** 783-858.

[20] 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.

[21] Massé, J. C. (1999). Asymptotics for the Tukey depth. Preprint.

[22] Mizera, I. (1998). On Depth and Deep Points: A Calculus. Preprint.

[23] Nolan, D. (1999). On min-max majority and deepest points. *Statist. Probab. Letters* **43** 325–334.

[24] Portnoy, S., and Mizera, I. (1999). Comment, as discussion of "Regression depth" by Rousseeuw and Hubert. *J. Amer. Statist. Assoc.* **94** 417-419

[25] Rousseeuw, P. J., and Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94** 388-402

[26] Serfling, R. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons.

[27] Serfling, R. (1984). Generalized L-, M-, and R-statistics. *Ann. Statist.* **12** 76–86.

[28] Sievers, G. L. (1978). Estimation of location: A large deviation comparison. *Ann. Statist.* **6** 610-618.

[29] Small, C. G. (1990). A survey of multidimensional medians. *Intern. Statist. Inst. Rev.* **58** 263–277.

[30] Tukey, J. W. (1975). Mathematics and picturing data. *Proc. Intern. Congr. Math. Vancouver 1974* **2** 523–531.

[31] Zuo, Y., and Serfling, R. (2000a). General notions of statistical depth function. To appear in *Ann. Statist.* **28** (2).

[32] Zuo, Y., and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. To appear in *Ann. Statist.* **28** (2).

[33] Zuo, Y., and Serfling, R. (2000c). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inference* **84** 55-79.

[34] Zuo, Y., and Serfling, R. (2000d). Nonparametric Notions of Multivariate "Scatter Measure" and "More Scattered" Based on Statistical Depth Functions. To appear in *J. Multiv. Analysis.*