Inequalities Relating Addition and Replacement Type Finite Sample Breakdown Points

Robert Serfling
Department of Mathematical Sciences
University of Texas at Dallas
Richardson, Texas 75083-0688, USA
Email: `serfling@utdallas.edu`
Website: `www.utdallas.edu/~serfling`

This paper is dedicated to Dr. A. K. Md. Ehsanes Saleh, in celebration of his distinguished accomplishments and contributions to statistical science, especially nonparametric methods.

Abstract

We explore relationships between the finite-sample addition breakdown point (ABP) and replacement breakdown point (RBP) of a statistic. Each concerns the minimum fraction of contaminants present in a sample, due to either addition or replacement, that can cause breakdown. Some authors prefer the ABP, which avoids the need to specify points to replace. Others argue the merits of the RBP, which avoids the conceptual issue of adding further points to the actual data. Zuo (2001) provides quantitative correspondences between the ABP and RBP when they depend only on the sample size and assume a particular form. In the present note we pursue their relationship in full generality, allowing dependence on data values and not restricting to any special form, thus including for example the Hodges-Lehmann location estimator and the sample halfspace median. We develop inequalities showing that the ABP and RBP are equivalent in the senses that (i) each corresponds to the other, through explicit expressions or by inequalities, although their values can slightly differ, and (ii) asymptotic limits, whether deterministic or almost sure, agree exactly. Therefore, as measures of robustness the ABP and RBP perform equivalently for practical purposes. This grants a pardon to authors who inadvertently commit the crime of comparing one estimator's ABP with another's RBP.

Key words and phrases: breakdown points; multivariate analysis; nonparametric methods; strong convergence.

AMS 2000 Subject Classification: Primary 62G99 Secondary 62H99.

1 Introduction and a Basic Relation

We explore relationships between the finite-sample addition breakdown point (ABP) and replacement breakdown point (RBP) of a statistic $T(\mathbb{X}_N)$, for $\mathbb{X}_N$ a data set of size N in $\mathbb{R}^d$ and $T(\mathbb{X}_N)$ some nonnegative real-valued measure of the robustness of some estimator of interest based on $\mathbb{X}_N$. For example, relative to a vector-valued location estimator $m(\mathbb{X}_N)$, we might choose $T_0(\mathbb{X}_N) = \|m(\mathbb{X}_N)\|$, where $\|\cdot\|$ denotes the usual Euclidean norm. For a positive definite matrix-valued scatter estimator $S(\mathbb{X}_N)$, we might take an appropriate function of the eigenvalues $\lambda_j(S(\mathbb{X}_N))$, $j = 1, \ldots, d$ (which are nonnegative in this case), for example, $T_1(\mathbb{X}_N) = \sum_{j=1}^d (\lambda_j(S(\mathbb{X}_N)) + \lambda_j(S(\mathbb{X}_N))^{-1})$. For simultaneous location and scatter estimation, $T(\mathbb{X}_N)$ may be the sum of $T_0(\mathbb{X}_N)$ and $T_1(\mathbb{X}_N)$. For data $\mathbb{X}_N$ relative to the linear regression model $Y = \sum_{i=1}^d \theta_i Z_i$ and $\hat{\boldsymbol{\theta}}(\mathbb{X}_N)$ some estimator of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, we take $T(\mathbb{X}_N) = \|\hat{\boldsymbol{\theta}}\|$. In such contexts, the stability of $T(\mathbb{X}_N)$ under corruption of sample values measures the robustness of the relevant estimator.

If for N held fixed $T(\mathbb{X}_N)$ can be taken to ∞ by introducing "contaminants" into the sample, then the relevant estimator is said to "break down", and a corresponding "breakdown point" is based on the minimal number of contaminants needed to produce this result. The notion of such a "finite sample" version of the asymptotic type breakdown point of Hampel (1968, 1971) was introduced by Donoho and Huber (1983). In the sequel we refer interchangeably to breakdown of $T(\mathbb{X}_N)$ and of the relevant estimator.

Two types of finite sample breakdown point have become popular, based on whether the contaminants are additions or replacements. *Addition breakdown* of $T(\mathbb{X}_N)$ occurs with k points $\mathbb{Y}_k^{(a)}$ added to the sample $\mathbb{X}_N$ if

$$\sup_{\mathbb{Y}_k^{(a)}} |T(\mathbb{X}_N) - T(\mathbb{X}_N, \mathbb{Y}_k)| = \infty, \tag{1}$$

where $T(\mathbb{X}_N, \mathbb{Y}_k)$ denotes the evaluation of $T(\cdot)$ over the combined sample $\{\mathbb{X}_N, \mathbb{Y}_k^{(a)}\}$ and the supremum is over all possible sets $\mathbb{Y}_k^{(a)}$ of k added points. Defining

$$k_A(T, N, \mathbb{X}_N) = \min\{k : T(\mathbb{X}_N) \text{ breaks down due to } k \text{ points added to } \mathbb{X}_N\},$$

the *addition breakdown point* of $T(\mathbb{X}_N)$ is then

$$ABP(T, N, \mathbb{X}_N) = \frac{k_A(T, N, \mathbb{X}_N)}{N + k_A(T, N, \mathbb{X}_N)}.$$

On the other hand, *replacement breakdown* of $T(\mathbb{X}_N)$ occurs with k points of $\mathbb{X}_N$ replaced by $\mathbb{Y}_k^{(r)}$ if

$$\max_{\alpha} \sup_{\mathbb{Y}_k^{(r)}} |T(\mathbb{X}_N) - T(\mathbb{X}_{N-k}^{(\alpha)}, \mathbb{Y}_k^{(r)})| = \infty, \tag{2}$$

where α indexes the $\binom{N}{k}$ subsets $\mathbb{X}_{N-k}^{(\alpha)}$ of size $N - k$ representing the possible sets of points in $\mathbb{X}_N$ *not* replaced. Defining

$$k_R(T, N, \mathbb{X}_N) = \min\{k : T(\mathbb{X}_N) \text{ breaks down due to } k \text{ replacements in } \mathbb{X}_N\},$$

the *replacement breakdown point* of $T(\mathbb{X}_N)$ is then

$$RBP(T, N) = \frac{k_R(T, N, \mathbb{X}_N)}{N}.$$

Some authors prefer the ABP, which often is somewhat easier to evaluate and often has a cleaner expression. For the ABP one needs not choose a set of points to be replaced, one merely adds in further points as one may please. This is especially helpful when the ABP and RBP depend upon the actual data values, for the RBP then depends on which set of points becomes replaced, among $\binom{N}{k}$ choices, for some choice of k. On the other hand, many authors (e.g., Rousseeuw and Leroy, 1987, pp. 117-118), argue the merits of the RBP, which involves just the single data set at hand and avoids possible conceptual issues associated with adding further observations to a given data set.

In order to establish some practical perspective on these two choices, Zuo (2001) provides quantitative correspondences between the ABP and the RBP in the case that

$$\textit{the ABP and RBP depend only on the sample size } N \tag{3}$$

and $k_A(T, N, \mathbb{X}_N) = k_A(T, N)$ satisfies

$$k_A(T, N, \mathbb{X}_N) = k_A(T, N) \textit{ has the form either } \lfloor aN + b \rfloor \textit{ or } \lceil aN + b \rceil, \tag{4}$$

for some constants a and b, where $\lfloor x \rfloor$ denotes the largest integer $\leq x$ and $\lceil \cdot \rceil$ denotes the smallest integer $\geq x$. Assumptions (3) and (4) are satisfied by many examples in the literature.

In the present note we pursue the relationship between addition and replacement breakdown points in full generality. Under (3), but without assuming (4), we develop inequalities for the ABP and RBP that yield asymptotic results and not only cover the case of (3) but also include examples such as the *Hodges-Lehmann location estimator*, which satisfies (3) but not (4). We also extend to the general case that (3) does not hold, i..e, that the ABP and RBP depend upon the actual data values in $\mathbb{X}_N$, thus allowing examples such as the *sample halfspace median*.

Despite numerical and conceptual differences, the ABP and RBP intuitively seem to be very similar ways to define the minimum fraction of contaminants in a sample that can cause breakdown of a statistic. We find that indeed the ABP and RBP are *equivalent*, in the senses that (i) each corresponds to the other, through explicit expressions or by inequalities, although their values can be slightly different, and (ii) their asymptotic limits, whether deterministic or almost sure, agree exactly. Therefore, *as measures of robustness of estimators, the ABP and RBP perform equivalently for practical purposes, giving the same value with negligible difference.* This grants a pardon to authors who inadvertently commit the crime of comparing one estimator's ABP with another's RBP.

We proceed as follows. Lemma 1 below gives a simple basic connection between addition and replacement breakdown, a result used in the sequel. Section 2 develops general results under assumption (3). Inequalities relating $k_A(T, N)$ and $k_R(T, N)$ are provided in Theorem 2 and inequalities relating the ABP and the RBP in Theorem 4. Further aspects are developed as well. Corollary 6, for example, asserts that the ABP and RBP have a common asymptotic limit. Example 9 treats the special case that $ABP(T, N) = N + m$ for an integer m, which satisfies both (3) and (4) and gathers into one convenient general form all of the examples treated in Zuo (2001). In Section 3, extensions to the general case not requiring (3) are carried out. This introduces the difficulty that the RBP involves many subsets of the given data set and the interdependence of these makes probabilistic analysis somewhat difficult. However, with the use of general inequalities that we establish, we obtain for the halfspace median, for example, that the uniform lower bound $1/(d+1)$ and (under symmetry conditions) the almost sure upper bound $1/3$, $d \geq 2$, derived by Donoho and Gasko (1992) and Chen (1995) for its ABP also apply to its RBP.

We conclude the present section with the following result giving a key connection between addition and replacement breakdown.

Lemma 1 *Replacement breakdown of $T(\mathbb{X}_N)$ with k points replaced is equivalent to addition breakdown of $T(\mathbb{X}_{N-k}^{(\alpha)})$ due to k points added for some $\alpha \in \{1, \ldots, \binom{N}{k}\}$.*

Proof. Note that (2) holds if and only if for some $\alpha \in \{1, \ldots, \binom{N}{k}\}$ we have

$$\sup_{\mathbb{Y}_k^{(r)}} |T(\mathbb{X}_N) - T(\mathbb{X}_{N-k}^{(\alpha)}, \mathbb{Y}_k^{(r)})| = \infty, \tag{5}$$

which by (1) is equivalent to *addition* breakdown of $T(\mathbb{X}_{N-k}^{(\alpha)})$ with k added points. □

2 Breakdown Points in a Special Case

Assume (3), which may be expressed as the assumption that k_A and k_R are well-defined and do not depend on the values of $\mathbb{X}_N$, i.e.,

$$k_A(T, N, \mathbb{X}_N) = k_A(T, N) \text{ and } k_R(T, N, \mathbb{X}_N) = k_R(T, N) \tag{6}$$

(possibly subject to some structural assumptions on the data, for example that $\mathbb{X}_N$ is in general position). An important implication of this assumption is that (2) holds if and only if

$$\sup_{\mathbb{Y}_k^{(r)}} |T(\mathbb{X}_N) - T(\mathbb{X}_{N-k}^{(\alpha)}, \mathbb{Y}_k^{(r)})| = \infty \tag{7}$$

holds for each $\alpha = 1, \ldots, \binom{N}{k}$. Hence Lemma 1 may be restated as

> *Replacement breakdown of $T(\mathbb{X}_N)$ with k points replaced is equivalent to addition breakdown of $T(\mathbb{X}_{N-k})$ due to k points added for any subset $\mathbb{X}_{N-k}$ of size $N-k$ in $\mathbb{X}_N$.*

For some results we will assume further that

$$k_A(T, N) \text{ is nondecreasing in } N \tag{8}$$

(which does not follow from the definitions but would be true for all typical choices of $T(\cdot)$). It is seen in Corollary 3 below that (8) implies the same property for k_R.

We now establish some productive inequalities regarding $k_A(T, N)$ and $k_R(T, N)$.

Theorem 2 *The functions k_A and k_R satisfy*

$$k_A(T, N - k_R(T, N)) \le k_R(T, N) \le k_A(T, N - k_R(T, N) + 1) \tag{9}$$

$$k_R(T, N + k_A(T, N)) \le k_A(T, N) \le k_R(T, N + k_A(T, N) - 1). \tag{10}$$

Further, under (8) we have equality in (10), i.e.,

$$k_R(T, N + k_A(T, N)) = k_A(T, N) = k_R(T, N + k_A(T, N) - 1). \tag{11}$$

Proof. (i) Replacement of $k_R(T,N)$ points of $\mathbb{X}_N$ yields breakdown of $T(\mathbb{X}_N)$. Hence, by Lemma 1, addition of $k_R(T,N)$ points to a subset $\mathbb{X}^{(\alpha)}_{N-k_R(T,N)}$ of size $N-k_R(T,N)$ in $\mathbb{X}_N$ causes breakdown of $T(\mathbb{X}^{(\alpha)}_{N-k_R(T,N)})$. Thus follows the first inequality of (9).

(ii) Replacement of any $k_R(T,N)-1$ points of $\mathbb{X}_N$ fails to yield breakdown of $T(\mathbb{X}_N)$. Hence, again by Lemma 1, addition of $k_R(T,N)-1$ points to any subset $\mathbb{X}^{(\alpha)}_{N-k_R(T,N)+1}$ of size $N-[k_R(T,N)-1]=N-k_R(T,N)+1$ cannot cause breakdown of $T(\mathbb{X}^{(\alpha)}_{N-k_R(T,N)+1})$, whereas addition of $k_A(N-k_R(T,N)+1)$ points does cause its breakdown. Thus $k_A(T,N-k_R(T,N)+1) > k_R(T,N)-1$ and the second inequality of (9) follows.

(iii) Addition of $k_A(T,N)$ points to $\mathbb{X}_N$ causes breakdown of $T(\mathbb{X}_N)$. Therefore, by Lemma 1, replacement of any $k_A(T,N)$ points in a sample $\mathbb{X}_{N+k_R(T,N)}$ of size $N+k_A(T,N)$ causes breakdown of $T(\mathbb{X}_{N+k_R(T,N)})$, giving the first inequality of (10).

(iv) Addition of $k_A(T,N)-1$ points to $\mathbb{X}_N$ fails to cause breakdown of $T(\mathbb{X}_N)$. Hence replacement of the $k_A(T,N)-1$ points of $\mathbb{X}_{N+k_A(T,N)-1}\backslash\mathbb{X}_N$ does not yield breakdown of $T(\mathbb{X}_{N+k_A(T,N)-1})$, and so $k_R(T,N+k_A(T,N)-1) > k_A(T,N)-1$ and the second inequality of (10) follows.

(v) Finally, suppose that (8) holds. Suppose also that

$$k_R(T,N) \text{ is nondecreasing in } N \tag{12}$$

does *not* hold for some N, i.e., $k_R(T,N) > k_R(T,N+1)$. Then, by (8) and (9), $k_R(T,N) \le k_A(T,N-k_R(T,N)+1) \le k_A(T,N-k_R(T,N+1)+1) \le k_R(T,N+1)$, a contradiction. Hence (12) does in fact hold. Then the extreme terms in (10) must be equal, yielding (11). □

Part (v) of the above proof yields

Corollary 3 *If $k_A(T,N)$ is nondecreasing in N, then so is $k_R(T,N)$.*

The above inequalities yield corresponding inequalities for breakdown points.

Theorem 4 *The breakdown points $ABP(T,N)$ and $RBP(T,N)$ satisfy*

$$ABP(T,N-k_R(T,N)) \le RBP(T,N) < ABP(T,N-k_R(T,N)+1)+\frac{1}{N} \tag{13}$$

$$RBP(T,N+k_A(T,N)) \le ABP(T,N) < RBP(T,N+k_A(T,N)-1). \tag{14}$$

Further, under (8) the first relation in (14) becomes equality, i.e.,

$$RBP(T,N+k_A(T,N)) = ABP(T,N) < RBP(T,N+k_A(T,N)-1). \tag{15}$$

Proof. (i) Using (9) we have

$$\begin{aligned} RBP(T,N) &= \frac{k_R(T,N)}{N} = \frac{k_R(T,N)}{[N-k_R(T,N)]+k_R(T,N)} \\ &\ge \frac{k_A(T,N-k_R(T,N))}{[N-k_R(T,N)]+k_A(T,N-k_R(T,N))} \\ &= ABP(T,N-k_R(T,N)) \end{aligned}$$

and

$$
\begin{aligned}
RBP(T,N) &= \frac{k_R(T,N)-1}{N} + \frac{1}{N} \\
&= \frac{k_R(T,N)-1}{[N-k_R(T,N)+1]+[k_R(T,N)-1]} + \frac{1}{N} \\
&\le \frac{k_A(T,N-k_R(T,N)+1)-1}{[N-k_R(T,N)+1]+[k_A(T,N-k_R(T,N)+1)-1]} + \frac{1}{N} \\
&< \frac{k_A(T,N-k_R(T,N)+1)}{[N-k_R(T,N)+1]+[k_A(T,N-k_R(T,N)+1)]} + \frac{1}{N} \\
&= ABP(T,N-k_R(T,N)+1) + \frac{1}{N},
\end{aligned}
$$

yielding (13).

(ii) Using (10) we have

$$
ABP(T,N) = \frac{k_A(T,N)}{N+k_A(T,N)} \ge \frac{k_R(T,N+k_A(T,N))}{N+k_A(T,N)} = RBP(T,N+k_A(T,N))
$$

and

$$
\begin{aligned}
ABP(T,N) &\le \frac{k_R(T,N+k_A(T,N)-1)}{N+k_A(T,N)} \\
&< \frac{k_R(T,N+k_A(T,N)-1)}{N+k_A(T,N)-1} = RBP(T,N+k_A(T,N)-1),
\end{aligned}
$$

yielding (14).

(iii) Finally, under (8) the "$\le$" and "$\ge$" in part (ii) each become "$=$", yielding (15). □

Remark 5 *The strict inequality $RBP(T,N+k_A(T,N)) < RBP(T,N+k_A(T,N)-1)$ implied by (14) and (15) may seem counterintuitive. However, when (11) holds, for example, this inequality follows immediately from the reverse inequality $N+k_A(T,N)-1 < N+k_A(T,N)$ satisfied by the denominators in their definitions. This is illustrated in Example 9 below.* □

A useful practical consequence of (13) and (14) together is that if either of ABP or RBP has a limit as $N \to \infty$, then the other has the same limit. Asymptotically, therefore, the ABP and RBP are interchangeable:

Corollary 6 $\lim_{N\to\infty} ABP(T,N) = \lim_{N\to\infty} RBP(T,N)$.

Proof. Let the ABP and RBP have respective limits L_A and L_R as $N \to \infty$. Then $k_A(T,N) \sim (L_A/(1-L_A))N$ and $k_R(T,N) \sim L_R N$, $N \to \infty$, and hence (9) yields $(L_A/(1-L_A))(1-L_R) = L_R$ and thus $L_A = L_R$. □

Remark 7 *Note that $ABP(T,N)$ is nondecreasing (nonincreasing) in N if and only if $k_A(T,N)/N$ is nondecreasing (nonincreasing) in N. The case that $k_A(T,N)/N$ is nondecreasing is a stronger assumption than (8).* □

A direct consequence of the fact in Remark 7 along with (13) is that when $k_A(T,N)/N$ is nondecreasing in N, then the ABP is approximately an upper bound to the RBP.

Corollary 8 *If $k_A(T,N)/N$ is nondecreasing in N, then so is ABP(T,N) and we have*

$$RBP(T,N) < ABP(T,N) + \frac{1}{N}. \tag{16}$$

We now illustrate, by a concrete example that covers many special cases, how the ABP and RBP can correspond explicitly. It suffices to show a correspondence between $k_A(T,N)$ and $k_R(T,N)$, and for the following example this is straightfoward by "solving" the simultaneous inequalities in Theorem 2.

Example 9 *If $k_A(T,N) = N+m$ for some integer m, then $k_R(T,N) = \lfloor\frac{N+m+1}{2}\rfloor$, and conversely.*
Proof. *(a) Suppose that $k_R(T,N) = \lfloor\frac{N+m+1}{2}\rfloor$. Then the two inequalities in (10) give*

$$\left\lfloor \frac{N + k_A(T,N) + m + 1}{2} \right\rfloor \le k_A(T,N) \le \left\lfloor \frac{N + k_A(T,N) + m}{2} \right\rfloor.$$

Using the fact that $k_A(\cdot)$ is integral, it follows immediately that

$$k_A(T,N) = \left\lfloor \frac{N + k_A(T,N) + m}{2} \right\rfloor,$$

yielding $k_A(T,N) = N+m$. (b) Conversely, if $k_A(T,N) = N+m$, then (9) gives

$$N - k_R(T,N) + m \le k_R(T,N) \le N - k_R(T,N) + m + 1,$$

and hence

$$\frac{N+m}{2} \le k_R(T,N) \le \frac{N+m+1}{2},$$

yielding that $k_R(T,N)$ equals either $\lfloor\frac{N+m}{2}\rfloor$ or $\lfloor\frac{N+m+1}{2}\rfloor$. By (a), we conclude that $k_R(T,N) = \lfloor\frac{N+m+1}{2}\rfloor$. □

Let us also illustrate the strict inequality discussed in Remark 5. We have

$$RBP(T, N + k_A(T,N)) = RBP(T, 2N+m) = \frac{\lfloor\frac{2N+2m+1}{2}\rfloor}{2N+m} = \frac{N+m}{2N+m},$$

and then by similar steps

$$RBP(T, N + k_A(T,N) - 1) = RBP(T, 2N+m-1) = \frac{N+m}{2N+m-1} > RBP(T, N + k_A(T,N)).$$

For $m = 0$, the ABP is exactly 1/2. Also, for any choice of m, we have 1/2 as the limit of both the ABP and the RBP,

$$\lim_{N\to\infty} ABP(T,N) = \lim_{N\to\infty} RBP(T,N) = 1/2.$$

Note that $k_A(N)/N$ in this example is monotone increasing in the (typical) case $m < 0$, in which case the ABP is an approximate upper bound to the RBP, i.e., (16) holds. □

Example 9 provides a convenient level of generality. All of the particular examples mentioned in Zuo (2001) may be conveniently gathered together as special cases of this one simple example, for various choices of $m \leq 0$, as follows.

(i) *The univariate median.* $m = 0$, $k_A(T,N) = N$, and $k_R(T,N) = \lfloor \frac{N+1}{2} \rfloor$.

(ii) *The spatial median in* $\mathbb{R}^d$. Again, $m = 0$, $k_A(T,N) = N$, and $k_R(T,N) = \lfloor \frac{N+1}{2} \rfloor$.

(iii) *The least median of squares estimator in the linear regression model* $Y = \sum_{i=1}^{d} \theta_i Z_i$, $d \geq 2$. $m = -2d+3$, $k_A(T,N) = N - 2d + 3$, and $k_R(T,N) = \lfloor \frac{N-2d+4}{2} \rfloor$.

(iv) *Simultaneous S-estimators of multivariate location and scatter* $m = -d$, $k_A(T,N) = N - d$, and $k_R(T,N) = \lfloor \frac{N-d+1}{2} \rfloor$.

For brief background discussion of these special cases, see Zuo (2001). The correspondence between the ABP and the RBP in Example 9 can be derived also from a somewhat more general but also more complicated structure covered in Theorems 2.1 and 2.2 of Zuo (2001), which themselves can also be proved using our inequalities.

Not all estimators with ABP satisfying (3) also satisfy (4), and not all robust estimators attain the highest asymptotic BP of 1/2. A typical example is the following.

Example 10 *The well-known univariate Hodges-Lehmann location estimator (Hodges and Lehmann, 1964) is simply the median of pairwise averages, median*$\{(X_i + X_j)/2\}$. *Extension to* $\mathbb{R}^d$ *using the spatial median and also considering* m*-wise averaging* ($m \geq 2$) *is carried out by Chaudhuri (1992). It is not difficult to argue, for the* m*-wise case, that*

$$k_R(T_m, N) = \max_{1 \leq j \leq N} \left\{ j : \frac{\binom{N-j}{m}}{\binom{N}{m}} \geq \frac{1}{2} \right\},$$

and that

$$RBP(T_m, \mathbb{X}_N) \to 1 - (1/2)^{1/m}, \ N \to \infty.$$

As for the corresponding ABP, adequate practical information is given by Corollary 6: it has the same limit as the RBP. □

3 Extension to the General Case

Now let $k_A(\cdot)$ and $k_R(\cdot)$ depend upon both N and the particular sample values of $\mathbb{X}_N$. For any M, denote by $\mathbb{X}_M$ the set of consecutive values $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M\}$ taken from the sequence $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots\}$ in $\mathbb{R}^d$. For some results we will assume the following analogue of (8):

$$k_A(T, N, \mathbb{X}_N) \leq k_A(T, N+1, \mathbb{X}_{N+1}). \tag{17}$$

We have the following analogue of Theorem 4.

Theorem 11 *The functions k_A and k_R satisfy*

$$\min_\alpha k_A(T, N - k_R(T, N, \mathbb{X}_N), \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)})$$

$$\leq k_R(T, N, \mathbb{X}_N) \tag{18}$$

$$\leq \min_\alpha k_A(T, N - k_R(T, N, \mathbb{X}_N) + 1, \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1}) \tag{19}$$

$$k_R(T, N + k_A(T, N, \mathbb{X}_N), \mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)})$$

$$\leq k_A(T, N, \mathbb{X}_N) \tag{20}$$

$$\leq k_R(T, N + k_A(T, N, \mathbb{X}_N) - 1, \mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)-1}). \tag{21}$$

Further, under (17) we have equality in (20) and (21).

Proof. (i) For some α, replacement of the $k_R(T, N, \mathbb{X}_N)$ points of the set $\mathbb{X}_N \backslash \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)}$ of size $N - k_R(T, N, \mathbb{X}_N)$ yields breakdown of $T(\mathbb{X}_N)$. Hence, by Lemma 1, addition of $k_R(T, N, \mathbb{X}_N)$ points to that particular $\mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)}$ causes breakdown of $T(\mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)})$. Thus follows (18).

(ii) Replacement of any $k_R(T, N, \mathbb{X}_N) - 1$ points of $\mathbb{X}_N$ fails to yield breakdown of $T(\mathbb{X}_N)$. Hence, again by Lemma 1, for each α, addition of $k_R(T, N, \mathbb{X}_N) - 1$ points to the subset $\mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1}$ of size $N - [k_R(T, N, \mathbb{X}_N) - 1] = N - k_R(T, N, \mathbb{X}_N) + 1$ cannot cause breakdown of $T(\mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1})$, whereas addition of $k_A(N - k_R(T, N, \mathbb{X}_N) + 1)$ points does cause its breakdown. Thus $k_A(T, N - k_R(T, N, \mathbb{X}_N) + 1, \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1}) > k_R(T, N, \mathbb{X}_N) - 1$ and (19) follows.

(iii) Addition of $k_A(T, N, \mathbb{X}_N)$ points to $\mathbb{X}_N$ causes breakdown of $T(\mathbb{X}_N)$. Therefore, by Lemma 1, replacement of the $k_A(T, N, \mathbb{X}_N)$ points of $\mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)} \backslash \mathbb{X}_N$ causes breakdown of $T(\mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)})$, giving (20).

(iv) Addition of $k_A(T, N, \mathbb{X}_N) - 1$ points to $\mathbb{X}_N$ fails to cause breakdown of $T(\mathbb{X}_N)$. Hence replacement of the $k_A(T, N, \mathbb{X}_N) - 1$ points of $\mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)-1} \backslash \mathbb{X}_N$ does not yield breakdown of $T(\mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)-1})$, and so $k_R(T, N + k_A(T, N, \mathbb{X}_N) - 1) > k_A(T, N, \mathbb{X}_N) - 1$ and (21) follows.

(v) Finally, suppose that (17) holds. Suppose also that

$$k_R(T, N, \mathbb{X}_N) > k_R(T, N + 1, \mathbb{X}_{N+1})$$

holds for some N. Then, using (17), (18), and (19), it is easily checked that

$$\begin{aligned} k_R(T, N, \mathbb{X}_N) &\leq \min_\alpha k_A(T, N - k_R(T, N, \mathbb{X}_N) + 1, \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1}) \\ &\leq \min_\beta k_A(T, N - k_R(T, N + 1, \mathbb{X}_{N+1}) + 1, \mathbb{X}^{(\beta)}_{N-k_R(T,N+1,\mathbb{X}_{N+1})+1}) \\ &\leq k_R(T, N + 1, \mathbb{X}_{N+1}), \end{aligned}$$

a contradiction. Thus, under (17), the relations in (20) and (21) become equality. □

Part (v) of the preceding proof yields the following analogue of Corollary 3.

Corollary 12 *If $k_A(T, N, \mathbb{X}_N)$ is nondecreasing in the sense of (17), then so is $k_R(T, N, \mathbb{X}_N)$.*

Theorem 11 yields inequalities for breakdown points, as an analogue of Theorem 4 proved by the same steps with obvious minor changes.

Theorem 13 *The breakdown points $ABP(T, N, \mathbb{X}_N)$ and $RBP(T, N, \mathbb{X}_N)$ satisfy*

$$\begin{aligned} \min_\alpha & ABP(T, N - k_R(T, N, \mathbb{X}_N), \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)}) \\ & \leq RBP(T, N, \mathbb{X}_N) & (22) \\ & < \min_\alpha ABP(T, N - k_R(T, N, \mathbb{X}_N) + 1, \mathbb{X}^{(\alpha)}_{N-k_R(T,N,\mathbb{X}_N)+1}) + \frac{1}{N} & (23) \end{aligned}$$

and

$$\begin{aligned} & RBP(T, N + k_A(T, N, \mathbb{X}_N), \mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)}) \\ & \quad \leq ABP(T, N, \mathbb{X}_N) & (24) \\ & \quad < RBP(T, N + k_A(T, N, \mathbb{X}_N) - 1, \mathbb{X}_{N+k_A(T,N,\mathbb{X}_N)-1}). & (25) \end{aligned}$$

Further, under (17) we have equality in (24).

While the ABP and RBP may depend upon the data values, they might under some assumptions on the parent model for $\mathbb{X}_N$ have deterministic limits in probability or almost surely as $N \to \infty$. If so, the inequalities of Theorem 13 show that these limits must agree.

Corollary 14 *If $ABP(T, N, \mathbb{X}_N)$ and $RBP(T, N, \mathbb{X}_N)$ have limits L_A and L_R, respectively, either in probability or almost surely, then these agree: $L_A = L_R$.*

Proof. If the ABP and RBP have limits L_A and L_R, then it is seen that $N - k_R(T, N, \mathbb{X}_N) + 1$ tends to ∞ in an appropriate sense and (23) yields

$$L_R \leq \lim_{N\to\infty} \left\{ ABP(T, N - k_R(T, N, \mathbb{X}_N) + 1, \mathbb{X}_{N-k_R(T,N,\mathbb{X}_N)+1}) + \frac{1}{N} \right\} = L_A.$$

A similar argument using (25) yields the opposite inequality. □

In some cases the ABP and RBP are subject to uniform bounds above or below, and the inequalities of Theorem 13 show that the bounds for one of these apply essentially unchanged to the other.

Corollary 15 *If $m_A \leq ABP(T, N, \mathbb{X}_N) \leq M_A$, then $m_A \leq RBP(T, N, \mathbb{X}_N) < M_A + 1/N$. If $m_R \leq RBP(T, N, \mathbb{X}_N) \leq M_R$, then $m_R \leq ABP(T, N, \mathbb{X}_N) < M_R$. In brief,*

$$\max\{m_A, m_R\} \quad \leq \quad ABP(T, N, \mathbb{X}_N) < \min\{M_A, M_R\}, \tag{26}$$

$$\max\{m_A, m_R\} \quad \leq \quad RBP(T, N, \mathbb{X}_N) < \min\{M_A + \frac{1}{N}, M_R\}. \tag{27}$$

Remark 16 *In the case that the ABP has a known limit L_A, but the situation for the RBP is unknown, we cannot apply Corollary 14. However, we conclude from Corollary 15 that $\limsup_{N\to\infty} RBP \leq L_A$.*

It is evident that Remark 7 applies with the extended $k_A(T, N, \mathbb{X}_N)$, and we readily obtain via (23) the following analogue of Corollary 8.

Corollary 17 *If $k_A(T, N, \mathbb{X}_N)/N$ is nondecreasing in the sense of (17), then $ABP(T, N, \mathbb{X}_N)$ is also nondecreasing in the same sense, and we have*

$$RBP(T, N, \mathbb{X}_N) < ABP(T, N, \mathbb{X}_N) + \frac{1}{N}. \quad (28)$$

Example 18 Halfspace Median. *Donoho and Gasko (1992) show that if $\mathbb{X}_N$ is in general position, then the ABP of the halfspace median is $\geq 1/(d+1)$, $d \geq 2$. Further, Donoho and Gasko (1992) and Chen (1995) show that if the underlying probability measure is absolutely continuous and angularly symmetric, then this ABP has almost sure limit L_A = 1/3, $N \to \infty$. From the preceding results, we thus conclude that*

$$RBP(T, N, \mathbb{X}_N) \geq \frac{1}{d+1}$$

for $d \geq 2$ and $\mathbb{X}_N$ is in general position, and that

$$\limsup_{N\to\infty} RBP(T, N, \mathbb{X}_N) \leq \frac{1}{3}$$

if the underlying probability measure is absolutely continuous and angularly symmetric. While the almost sure upper bound of 1/3 for the RBP is sufficient information for practical purposes, we conjecture that in fact the almost sure limit L_R exists and = 1/3. For some reinforcement of this conjecture, we note a useful empirical illustration of Zuo (2003, Figures 3 and 4), which shows that in a sample of size 20 from the standard normal distribution, the halfspace median can resist replacement of 6 points by outliers but not replacement of 7 points. □

Acknowledgment

Support by NSF Grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

References

[1] Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Annals of Statistics* **20** 897–916.

[2] Chen, Z. (1995). Bounds for the breakdown point of the simplicial median. *Journal of Multivariate Analysis* **55** 1–13.

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

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

[5] Hampel, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley.

[6] Hampel, F. R. (1971). A general qualitative deinition of robustness. *Annals of Mathematical Statistics* **42** 1887–1896.

[7] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.

[8] Zuo, Y. (2001). Some quantitative relationships between two types of finite sample breakdown point. *Statistics & Probability Letters* **51** 369–375.

[9] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.