REVSTAT – Statistical Journal
Volume 5, Number 1, March 2007, 1–17

THE BREAKDOWN POINT — EXAMPLES AND COUNTEREXAMPLES

Authors: P.L. Davies
– University of Duisburg-Essen, Germany, and
Technical University Eindhoven, Netherlands
laurie.davies@uni-essen.de

U. Gather
– University of Dortmund, Germany
gather@statistik.uni-dortmund.de

Abstract:

- The breakdown point plays an important though at times controversial role in statistics. In situations in which it has proved most successful there is a group of transformations which act on the sample space and which give rise to an equivariance structure. For equivariant functionals, that is those functionals which respect the group structure, a non-trivial upper bound for the breakdown point was derived in Davies and Gather (2005). The present paper briefly repeats the main results of Davies and Gather (2005) but is mainly concerned with giving additional insight into the concept of breakdown point. In particular, we discuss the attainability of the bound and the dependence of the breakdown point on the sample or distribution and on the metrics used in its definition.

Key-Words:

- *equivariance; breakdown point; robust statistics.*

AMS Subject Classification:

- Primary: 62G07; Secondary: 65D10, 62G20.

1. INTRODUCTION

The breakdown point is one of the most popular measures of robustness of a statistical procedure. Originally introduced for location functionals (Hampel, 1968, 1971) the concept has been generalized to scale, regression and — with more or less success — to other situations.

In Huber's functional analytic approach to robustness breakdown is related to the boundedness of a functional and the breakdown point is defined in terms of the sizes of neighbourhoods on the space of distributions. A simple and intuitive definition of the breakdown point but one restricted to finite samples, the finite sample breakdown point, was introduced by Donoho (1982) and Donoho and Huber (1983). Successful applications of the concept of breakdown point have been to the location, scale and regression models in $\mathbb{R}^k$ and to models which are intimately related to these (see for example Ellis and Morgenthaler, 1992, Davies and Gather, 1993, Hubert, 1997, Terbeck and Davies, 1998, He and Fung, 2000, Müller and Uhlig, 2001). The reason for this is that such models have a rich equivariance structure deriving from the translation or affine group operating on $\mathbb{R}^k$. By restricting the class of statistical functionals to those with the appropriate equivariance structure one can prove the existence of non-trivial highest breakdown points (Davies and Gather, 2005), which in many cases can be achieved, at least locally (Huber, 1981, Davies, 1993).

It is the aim of this paper to provide some additional insight into the definition of the breakdown point, to point out the limits of the concept and to give some results on the attainment of the upper bound.

We proceed as follows: Chapter 2 summarizes the definitions and theorems of Davies and Gather (2005). Chapter 3 shows via examples that the breakdown point is a local concept. Chapter 4 is devoted to the attainability of the bound and Chapter 5 to the choice of metrics. Chapter 6 contains some concluding remarks.

2. DEFINITIONS AND BOUNDS FOR THE BREAKDOWN POINT

Let T be a functional defined on some subfamily $\mathcal{P}_T$ of the family $\mathcal{P}$ of all distributions on a sample space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ which takes its values in some metric space (Θ, D) with

$$\sup_{\theta_1,\theta_2\in\Theta} D(\theta_1,\theta_2) = \infty \ . \tag{2.1}$$

The finite sample breakdown point of T at a sample $\boldsymbol{x}_n = (x_1, ..., x_n)$, $x_i \in \mathcal{X}$, $i = 1, ..., n$, is defined as

$$\text{fsbp}\,(T, \boldsymbol{x}_n, D) \;=\; \frac{1}{n}\,\min\Big\{k \in \{1, ..., n\}\colon \sup_{Q_{n,k}} D\big(T(P_n), T(Q_{n,k})\big) = \infty\Big\} \tag{2.2}$$

where $P_n = \sum_{i=1}^n \delta_{x_i}/n$ and $Q_{n,k}$ is the empirical distribution of a replacement sample with at least $n-k$ points from the original sample $\boldsymbol{x}_n$.

Example 2.1. If T is the median functional T_{med} defined on $\mathcal{P}_T = \mathcal{P}$ with $\Theta = \mathbb{R}$, and $D(\theta_1, \theta_2) = |\theta_1 - \theta_2|$, then

$$\text{fsbp}\,(T_{\text{med}}, x, D) \;=\; \left\lfloor \frac{n+1}{2} \right\rfloor / \,n\ . \tag{2.3}$$

A distributional definition of the breakdown point requires a metric d on $\mathcal{P}$ with

$$\sup_{P,Q\in\mathcal{P}} d(P, Q) \;=\; 1\ .$$

The breakdown point of a functional T at a distribution $P \in \mathcal{P}_T$ w.r.t. d and D is then defined by

$$\epsilon^*(T, P, d, D) \;=\; \inf\Big\{\epsilon > 0\colon \sup_{d(P,Q)<\epsilon} D\big(T(P), T(Q)\big) = \infty\Big\} \tag{2.4}$$

where $D\big(T(P), T(Q)\big) := \infty$ if $Q \notin \mathcal{P}_T$.

Example 2.2. Let $\mathcal{P}$ and D be as in Example 2.1 and d be the Kolmogorov-metric $d_k(P, Q) = \sup_x |F_P(x) - F_Q(x)|$. For the expectation functional T_E

$$T_E(P) \;=\; E(P) \;:=\! \int x\, dP(x)\ , \qquad \mathcal{P}_T \;=\; \Big\{P \in \mathcal{P}\colon\ E(P) \text{ exists}\Big\}$$

we have $\epsilon^*(T_E, P, d, D) = 0$ for any $P \in \mathcal{P}_T$, in contrast to the median for which $\epsilon^*(T_{\text{med}}, P, d, D) = 1/2$.

As already pointed out in the introduction the derivation of a non-trivial upper bound for the breakdown point requires a group structure. Assume that G is a group of measurable transformations of the sample space $\mathcal{X}$ onto itself. Then G induces a group of transformations of $\mathcal{P}$ onto itself via $P^g(B) = P(g^{-1}(B))$ for all sets $B \in \mathcal{B}(\mathcal{X})$. Let $H_g = \{h_g\colon g \in G\}$ be the group of transformations $h_g\colon \Theta \to \Theta$ which describes the equivariance structure of the problem. A functional $T\colon \mathcal{P}_T \to \Theta$ is called equivariant with respect to G if and only if $\mathcal{P}_T$ is closed under G and

$$T(P^g) = h_g\big(T(P)\big) \qquad \text{for all}\quad g \in G,\ \ P \in \mathcal{P}_T\ . \tag{2.5}$$

Let

$$G_1 := \Big\{ g \in G\colon \lim_{n\to\infty} \inf_{\theta\in\Theta} D\big(\theta, h_{g^n}(\theta)\big) = \infty \Big\} \tag{2.6}$$

and define

$$\Delta(Q) := \sup\Big\{ Q(B)\colon\ B \in \mathcal{B}(\mathcal{X}),\ g_{|B} = \iota_{|B} \text{ for some } g \in G_1 \Big\} \tag{2.7}$$

where ι is the unit element of G. We cite the main result from Davies and Gather (2005):

Theorem 2.1. *Suppose that the metrics d and D satisfy the properties given above and additionally*

$$d\Big(\alpha P + (1-\alpha) Q_1,\ \alpha P + (1-\alpha) Q_2\Big) \le 1-\alpha\,, \qquad P, Q_1, Q_2 \in \mathcal{P},\ 0<\alpha<1\,, \tag{2.8}$$

$$G_1 \neq \emptyset\,. \tag{2.9}$$

Then for all G-equivariant functionals $T\colon \mathcal{P}_T \to \Theta$, for all $P \in \mathcal{P}_T$ and for all $\boldsymbol{x}_n$ we have respectively

a) $\epsilon^*(T, P, d, D) \le \dfrac{\big(1-\Delta(P)\big)}{2}$,

b) $\mathrm{fsbp}(T, \boldsymbol{x}_n, D) \le \left\lfloor \dfrac{n - n\,\Delta(P_n) + 1}{2} \right\rfloor \Big/ n\,.$

Proof: **a**) cf. Davies and Gather (2005).

b) The proof is similar to a) but it is not given in Davies and Gather (2005). We present it here as it illustrates the simplicity of the idea of the finite sample breakdown point. The basic idea of all such proofs may be found in Huber (1981) although it was clearly known to Hampel (1975) who stated the breakdown point of what is now known as the LMS estimator (see Rousseeuw, 1984). Donoho and Huber (1983) give the first calculations for the finite sample breakdown point both for multivariate location and for a high breakdown linear regression estimator based on the multivariate location estimator of Donoho (1982). The corresponding calculations for the LMS estimator may be found in Rousseeuw (1984). Firstly we note that there are exactly $n\,\Delta(P_n)$ points in $\boldsymbol{x}_n$ for which $g(x_i) = x_i$ for some $g \in G_1$. We assume without loss of generality that these are the sample points $x_1, ..., x_{n\Delta(P_n)}$. If $\Delta(P_n) = 0$ there are no such points and some obvious alterations to the following proof are required. To ease the notation we write

$$l(n) = \left\lfloor \frac{n - n\,\Delta(P_n) + 1}{2} \right\rfloor .$$

We consider the sample $\boldsymbol{y}^*_{n,k}$ given by

$$\boldsymbol{y}^*_{n,k} = \Big(x_1, ..., x_{n\Delta(P_n)}, ..., x_{n-l(n)},\ g^m(x_{n-l(n)+1}), ..., g^m(x_n) \Big)$$

for some $m \ge 1$ and some $g \in G_1$. We denote its empirical distribution by $Q^*_{n,k}$.

The sample $\boldsymbol{y}^*_{n,k}$ contains at least $n - l(n)$ points of the original sample $\boldsymbol{x}_n$. The transformed sample $g^{-m}(\boldsymbol{y}^*_{n,k})$ is equal to

$$\Big(x_1, ..., x_{n\Delta(P_n)},\ g^{-m}(x_{n\Delta(P_n)+1}), ..., g^{-m}(x_{n-l(n)}),\ x_{n-l(n)+1}, ..., x_n\Big)\ .$$

It contains at least $n\,\Delta(P_n) + l(n)$ points of the original sample $\boldsymbol{x}_n$ and as

$$n\,\Delta(P_n) + l(n)\ \geq\ n - l(n)$$

it contains at least $n - l(n)$ points of $\boldsymbol{x}_n$. By the equivariance of T we have

$$T(Q^{*g^{-m}}_{n,k})\ =\ h_{g^{-m}}\big(T(Q^*_{n,k})\big)$$

from which it follows

$$D\Big(h_{g^{-m}}\big(T(Q^*_{n,k})\big), T(Q^*_{n,k})\Big)\ \leq\ D\big(T(P_n), T(Q^*_{n,k})\big) + D\big(T(P_n), T(Q^{*g^{-m}}_{n,k})\big)\ .$$

From $\lim_{n\to\infty} \inf_{\theta} D(\theta, h_{g^n}(\theta)) = \infty$ for all $g \in G_1$ we have

$$\lim_{m\to\infty} D\Big(h_{g^{-m}}\big(T(Q^*_{n,k})\big), T(Q^*_{n,k})\Big)\ =\ \infty$$

and hence $D\big(T(P_n), T(Q^*_{n,k})\big)$ and $D\big(T(P_n), T(Q^{*g^{-m}}_{n,k})\big)$ cannot both remain bounded. We conclude that for any $k \geq \lfloor \frac{n - n\Delta(P_n)+1}{2} \rfloor$

$$\sup_{Q_{n,k}} D\big(T(P_n), T(Q_{n,k})\big)\ =\ \infty$$

from which the claim of the theorem follows. □

For examples of Theorem 2.1 we refer to Davies and Gather (2005).

3. THE BREAKDOWN POINT IS A LOCAL CONCEPT

As seen above the median T_{med} has a finite sample breakdown point of $\lfloor(n+1)/2\rfloor$ at every real sample $\boldsymbol{x}_n$ and this is the highest possible value for translation equivariant location functionals. If we consider scale functionals then the situation is somewhat different. The statistical folklore is that the highest possible finite sample breakdown point for any affine equivariant scale functional is $\lfloor n/2\rfloor/n$ and that this is attained by the median absolute deviation functional T_{MAD}. Some authors (Croux and Rousseeuw, 1992, Davies, 1993) are aware that this is not correct as is shown by the following sample

(3.1) $$\boldsymbol{x}_{11} = \big(1.0, 1.8, 1.3, 1.3, 1.9, 1.1, 1.3, 1.6, 1.7, 1.3, 1.3\big)\ .$$

The fsbp of T_{MAD} at this sample is 1/11. This can be seen by replacing the data point 1.0 by 1.3 so that for the altered data set $T_{\text{MAD}}=0$ which is conventionally defined as breakdown. If a sample has no repeated observations then T_{MAD} has a finite sample breakdown point of $\lfloor n/2 \rfloor / n$ and this is indeed the highest possible finite sample breakdown point for a scale functional. The difference between the maximal finite sample breakdown points for location and scale functionals is explained by Theorem 2.1. For the sample (3.1) we have $\Delta(P_n)=5/11$ and the theorem gives

$$\text{fsbp}\big(T_{\text{MAD}}, \boldsymbol{x}_{11}, D\big) \;\le\; 3/11\ .$$

For a sample $\tilde{\boldsymbol{x}}_{11}$ without ties we have $\Delta(P_n)=1/n$ and the theorem yields

$$\text{fsbp}\big(T_{\text{MAD}}, \tilde{\boldsymbol{x}}_{11}, D\big) \;\le\; \left\lfloor \frac{n}{2} \right\rfloor / n \;=\; 5/11\ .$$

From the above it follows that T_{MAD} may or may not attain the upper bound. We study this in more detail in the next chapter.

4. ATTAINING THE BOUND

4.1. Location functionals

From Theorem 2.1 above it is clear that the maximum breakdown point for translation equivariant location functionals is 1/2. This bound is sharp as is shown by the location equivariant L_1-functional

$$T(P) \;=\; \text{argmin}_\mu \int \big(\|x-\mu\| - \|x\|\big)\, dP(x)\ . \tag{4.1}$$

In general the L_1-functional is not regarded as a satisfactory location functional as it is not affine equivariant in dimensions higher than one. For an affinely equivariant location functional the set G_1 of (2.6) is now the set of pure non-zero translations and it follows that $\Delta(P)=0$ for any distribution P. Theorem 2.1 gives an upper bound of 1/2 which is clearly attainable in one dimension. It is not however clear whether this bound is attainable in higher dimensions. Work has been done in this direction but it is not conclusive (Rousseeuw and Leroy, 1987, Niinimaa, Oja and Tableman, 1990, Lopuhaä and Rousseeuw, 1991, Gordaliza, 1991, Lopuhaä, 1992, Donoho and Gasko, 1992, Davies and Gather, 2005, Chapter 5 and the Discussion of Rousseeuw in Davies and Gather, 2005).

We first point out that the bound 1/2 is not globally sharp. Take a discrete measure in $\mathbb{R}^2$ with point mass 1/3 on the points $x_1=(0,1)$, $x_2=(0,-1)$, $x_3=(\sqrt{3},0)$. The points form a regular simplex. For symmetry reasons every affinely equivariant location functional must yield the value $(1/\sqrt{3},0)$. On replacing $(\sqrt{3},0)$ by $(\eta\sqrt{3},0)$ it is clear that each affinely equivariant location functional must result in $(\eta/\sqrt{3},0)$. On letting $\eta \to \infty$ it follows that the breakdown point of every affinely equivariant location functional cannot exceed 1/3. In k dimensions one can prove in a similar manner that $1/(k+1)$ is the maximal breakdown point for points on a regular simplex with $k+1$ sides.

In spite of the above example we now show that there are probability distributions at which the finite sample replacement breakdown point is 1/2 even if this cannot be obtained globally. We consider a sample $\boldsymbol{x}_n = (x_1, ..., x_n)$ of size n in $\mathbb{R}^k$ and form the empirical measure P_n given by $P_n = 1/n \sum_{i=1}^n \delta_{x_i}$. To obtain our goal we define an appropriate affinely equivariant location functional T at P_n^A for all affine transformations A and also at all measures of the form P_n^{*A}. Here P_n^* is any empirical measure obtained from $\boldsymbol{x}_n$ by altering the values of at most $\lfloor (n-1)/2 \rfloor$ of the x_i. The new sample will be denoted by $\boldsymbol{x}_n^* = (x_1^*, ..., x_n^*)$. We have to show that the values of $T(P_n^{*A})$ can be defined in such a way that

$$T(P_n^A) = A\big(T(P_n)\big) \;, \tag{4.2}$$

$$T(P_n^{*A}) = A\big(T(P_n^*)\big) \tag{4.3}$$

and

$$\sup_{P_n^*} \big|T(P_n) - T(P_n^*)\big| < \infty \;. \tag{4.4}$$

This is done in Appendix A.

We note that the Sample conditions 1 and 2 in Appendix A are satisfied for an i.i.d. Gaussian sample of size n if n is sufficiently large. We indicate how this may be shown in Appendix B.

4.2. Scatter functionals

At the sample (3.1) above the median absolute deviation T_{MAD} has a finite sample breakdown point of 1/11 compared with the upper bound of 3/11 given by Theorem 2.1. We consider a modification of T_{MAD} as defined in Davies and Gather (2005) which attains the upper bound.

For a probability measure P the interval $I(P,\lambda)$ is defined by

$$I(P,\lambda) = \Big[\text{med}(P) - \lambda,\ \text{med}(P) + \lambda\Big] \;.$$

We write

$$\Delta(P,\lambda) = \max\Big\{P(\{x\})\colon\ x \in I(P,\lambda)\Big\} \;.$$

The new scale functional T^*_{MAD} is defined by

$$T^*_{\text{MAD}}(P) = \min\Big\{\lambda\colon\ P\big(I(P,\lambda)\big) \ge \big(1 + \Delta(P,\lambda)\big)/2\Big\} \;.$$

We shall show

$$\text{fsbp}\big(T^*_{\text{MAD}}, \boldsymbol{x}_n, D\big) = \left\lfloor \frac{n - n\,\Delta(P_n) + 1}{2} \right\rfloor / n \; . \tag{4.5}$$

We consider a replacement sample $\boldsymbol{x}'_n$ with

$$n_1 + n_2 = m < \Big\lfloor \big(n - n\,\Delta(P_n) + 1\big)/2 \Big\rfloor$$

points replaced and with empirical distribution P'_n. We show firstly that $T^*_{\text{MAD}}(P'_n)$ does not explode. Let λ' be such that the interval $\big[\text{med}(P'_n) - \lambda', \text{med}(P'_n) + \lambda'\big]$ contains the original sample $\boldsymbol{x}_n$. As the median does not explode we see that λ' remains bounded over all replacement samples. Clearly if $T^*_{\text{MAD}}(P'_n)$ is to explode $\boldsymbol{x}'_n$ must contain points outside of this interval. We denote the number of such points by n_1. We use n_2 points to increase the size of the largest atom of $\boldsymbol{x}'_n$ in the interval. This is clearly done by placing these points at the largest atom of $\boldsymbol{x}_n$. The size of the largest atom of $\boldsymbol{x}'_n$ in the interval is therefore at most $\Delta(P_n) + n_2/n$. It follows that $T^*_{\text{MAD}}(P'_n) \le \lambda'$ if the interval contains at least $\big(n + n\,\Delta(P_n) + n_2\big)/2$ observations. This will be the case if $n - n_1 \ge \big(n + n\,\Delta(P_n) + n_2\big)/2$ which reduces to $n_1 + n_2/2 \le n\big(1 - \Delta(P_n)\big)/2$ which holds as

$$n_1 + n_2/2 \le n_1 + n_2 < \Big\lfloor n\big(1 - \Delta(P_n) + 1\big) \Big\rfloor / 2 \; .$$

It remains to show that $T^*_{\text{MAD}}(P'_n)$ does not implode to zero. For this to happen we would have to be able to construct a replacement sample for which the interval $I(P', \lambda)$ is arbitrarily small but for which $P'\big(I(P', \lambda)\big) \ge \big(1 + \Delta(P', \lambda)\big)/2$. In order for the interval to be arbitrarily small it must contain either no points of the original sample $\boldsymbol{x}_n$ or just one atom. In the latter case we denote the size of the atom by $\Delta_1(P_n)$. Suppose we replace $n_1 + n_2$ points and that the n_2 points form the largest atom in the interval $I(P', \lambda)$. We see that if $n_2 \ge n\,\Delta_1(P_n)$ then

$$n_1 + n_2 + n\,\Delta_1(P_n) \ge (n + n_2)/2$$

which implies

$$2n_1 + 2n_2 \ge 2n_1 + n_2 + n\,\Delta_1(P_n) \ge n > n - n\,\Delta(P_n)$$

which contradicts $n_1 + n_2 < \big\lfloor n\big(1 - \Delta(P_n) + 1\big) \big\rfloor / 2$. If the n_2 replacement points do not compose the largest atom then this must be of size at least $\Delta_1(P_n)$ which implies

$$n_1 + n_2 + n\,\Delta_1(P_n) \ge \big(n + n\,\Delta_1(P_n)\big)/2$$

and hence

$$2n_1 + 2n_2 \ge n - n\,\Delta_1(P_n) \ge n - n\,\Delta(P_n)$$

which again contradicts $n_1 + n_2 < \big\lfloor n\big(1 - \Delta(P_n) + 1\big) \big\rfloor / 2$. We conclude that $T^*_{\text{MAD}}(P'_n)$ cannot implode, and thus (4.5) is shown.

5. THE CHOICE OF THE METRICS d AND D

5.1. The metric d

Considering the parts a) and b) of Theorem 2.1 we note that there is in fact a direct connection between the two results. We consider the total variation metric d_{tv} defined by

$$d_{tv}(P,Q) \;=\; \sup_{B\in\mathcal{B}(\mathcal{X})} \bigl|P(B)-Q(B)\bigr| \;.$$

If $\mathcal{B}(\mathcal{X})$ "shatters" every finite set of points in $\mathcal{X}$ then

$$d_{tv}(P_n,P_n^*) \;=\; k/n$$

where P_n denotes the empirical measure deriving from $(x_1,...,x_n)$ and P_n^* that deriving from $(x_1^*,...,x_n^*)$ with the two samples differing in exactly k points. Suppose now that $\epsilon^*(T,P_n,d_{tv},D) = \bigl(1-\Delta(P_n)\bigr)/2$. If $k < n\bigl(1-\Delta(P_n)\bigr)/2$ then breakdown in the sense of finite sample breakdown point cannot occur and we see that

$$\text{(5.1)} \qquad \operatorname{fsbp}\bigl(T,\boldsymbol{x}_n,D\bigr) \;\geq\; \left\lfloor \frac{n-n\,\Delta(P_n)}{2} \right\rfloor \Big/ n \;.$$

Unfortunately the inequality of Theorem 2.1 b) seems not to be provable in the same manner.

We point out that the breakdown point is not necessarily the same for all metrics d. A simple counterexample is provided by the scale problem in $\mathbb{R}$. If we use the Kolmogorov metric then the breakdown point of T_{MAD} at an atomless distribution is 1/4 (Huber, 1981, page 110). However if we use the Kuiper metric d_{ku}^1 defined in (5.3) below then the breakdown point is 1/2 in spite of the fact that both metrics satisfy the conditions of the theorem. More generally if d' and d'' are two metrics satisfying $\sup_{P,Q\in\mathcal{P}} d(P,Q) = 1$ and (2.8) and such that $d' \leq d''$ then

$$\text{(5.2)} \qquad \epsilon^*(T,P,d',D) \;\leq\; \epsilon^*(T,P,d'',D) \;\leq\; \bigl(1-\Delta(P)\bigr)/2 \;.$$

In particular if $\epsilon^*(T,P,d',D) = \bigl(1-\Delta(P)\bigr)/2$ then $\epsilon^*(T,P,d'',D) = \bigl(1-\Delta(P)\bigr)/2$. A class of ordered metrics is provided by the generalized Kuiper metrics d_{ku}^m defined by

$$\text{(5.3)} \qquad d_{ku}^m(P,Q) \;=\; \sup\left\{ \left| \sum_{k=1}^{m} \bigl(P(I_k)-Q(I_k)\bigr) \right| :\; I_1,...,I_m \text{ disjoint intervals} \right\}.$$

We have

$$\text{(5.4)} \qquad d_{ku}^1 \leq \;...\; \leq d_{ku}^m \;.$$

For further remarks on the choice of d we refer to Davies and Gather (2005), Rejoinder and for a related but different generalization of the Kuiper metric of use in the context of the modality of densities we refer to Davies and Kovac (2004).

5.2. The metric D

As we have seen in the case of d above there seems to be no canonical choice: different choices of d can lead to different breakdown points. A similar problem exists with respect to the metric D on Θ. In the discussion of Tyler in Davies and Gather (2005) it was also pointed out that it might be difficult to achieve (2.1) when Θ is a compact space. This problem is discussed in the Rejoinder of Davies and Gather (2005), Chapter 6, and solved in Davies and Gather (2006) with applications to directional data.

We now indicate a possibility of making D dependent on d. The idea is that two parameter values θ_1 and θ_2 are far apart with respect to D if and only if the corresponding distributions are far apart with respect to d. We illustrate the idea using the location problem in $\mathbb{R}$. Suppose we have data with empirical distribution P_n and two values of the location parameter θ_1 and θ_2. We transform the data using the translations θ_1 and θ_2 which gives rise to two further distributions $P_n(\cdot - \theta_1)$ and $P_n(\cdot - \theta_2)$. If these two distributions are clearly distinguishable then $d\big(P_n(\cdot - \theta_1), P_n(\cdot - \theta_2)\big)$ will be almost one. An opposed case is provided by an autoregressive process of order one. The parameter space is $\Theta = (-1, 1)$ and this may be metricized in such a manner that $D(\theta_1, \theta_2)$ tends to infinity for fixed θ_1 as θ_2 tends to the boundary. However values of θ close to, on, or even beyond the boundary, may not be empirically distinguishable from values of θ in the parameter space. A sample of size $n = 100$ generated with $\theta_1 = 0.95$ is not easily distinguishable from a series generated with $\theta_2 = 0.9999$ even though $D(\theta_1, \theta_2)$ is large.

We now give a choice of D in terms of d and such that (2.1) is satisfied. We set

$$G(\theta_1, \theta_2) \;=\; \Big\{g \in G\colon\; h_g(\theta_1) = \theta_2\Big\}$$

and then define D by

$$D(\theta_1, \theta_2) \;=\; D_P(\theta_1, \theta_2) \;=\inf_{g \in G(\theta_1, \theta_2)} \big|\log\big(1 - d(P^g, P)\big)\big| \;. \tag{5.5}$$

The interpretation is that we associate P with the parameter value θ_1 and P^g with the parameter value θ_2. The requirement (2.1) will only hold if $d(P^g, P)$ may be arbitrarily close to one so that the distributions associated with θ_1 and θ_2 are as far apart as possible. It is easily checked that D defines a pseudometric on Θ, which is sufficient for our purposes; namely $D_P \geq 0$, D_P is symmetric and satisfies the triangle inequality. In some situations it seems reasonable to require that d and D be invariant with respect to the groups G and H_G respectively. If d is G-invariant, i.e.

$$d(P,Q) = d(P^g, Q^g), \qquad \text{for all} \quad P, Q \in \mathcal{P}, \quad g \in G \ ,$$

then D, defined by (5.5), inherits the invariance, i.e.

$$D(\theta_1, \theta_2) = D\big(h_g(\theta_1), h_g(\theta_2)\big), \qquad \text{for all} \quad \theta_1, \theta_2 \in \Theta, \quad g \in G \ .$$

The G-invariance of d can often be met.

6. FINAL REMARKS

We conclude with a small graphic showing the connections between all ingredients which are necessary for a meaningful breakdown point concept.

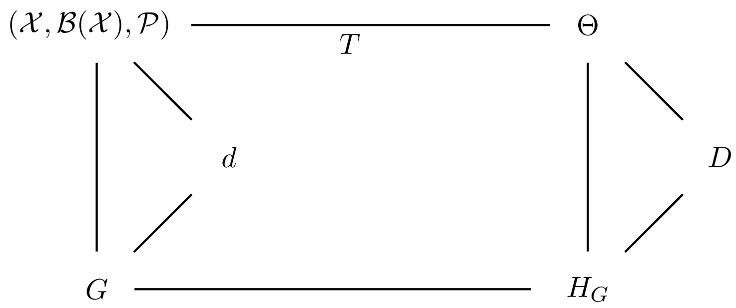

Figure 1: Connections.

We point out that each object in this graphic has an important influence on the breakdown point and its upper bound:

- $\epsilon^*(T, P, d, D)$ depends on P as shown in Chapter 3, and it depends on the metrics d and D as discussed in Chapter 5.
- It is the equivariance structure w.r.t. the group G which allows to prove an upper bound for $\epsilon^*(T, P, d, D)$ and it is the condition $G_1 \neq \emptyset$ which provides the main step in the proof. In particular, the choice of the group G determines $\Delta(P)$, thereby the upper bound, as well as its attainability. For many $\mathcal{P}$, T and G the attainability of the bound remains an open problem.

APPENDIX A

We consider the constraints imposed upon us when defining $T(P_n^*)$. We start with the internal constraints which apply to each P_n^* without reference to the other measures.

- **Case 1:** $P_n^{*A_1} \neq P_n^{*A_2}$ for any two different affine transformations A_1 and A_2. This is seen to reduce to $P_n^{*A} \neq P_n^*$ for any affine transformation A which is not the identity. If this is the case then there are no restrictions on the choice of $T(P_n^*)$. Having chosen it we extend the definition of T to all the measures P_n^{*A} by $T(P_n^{*A}) = A\big(T(P_n^*)\big)$.

- **Case 2:** $P_n^{*A} = P_n^*$ for some affine transformation A which is not the identity. If this is the case then A is unique and there exists a permutation π of $\{1, ..., n\}$ such that $A(x_i) = x_{\pi(i)}$. This implies that for each i we can form cycles
$$\Big(x_i,\, A(x_i),\, ...,\, A^{m_i-1}(x_i)\Big)$$
with $A^{m_i}(x_i) = x_i$. From this we see that for some sufficiently large m $A^m(x_i) = x_i$ for all i. On writing $A(x) = \alpha(x) + a$ we see that if the x_i, $i = 1, ..., n$, span $\mathbb{R}^k$ then $\alpha^m = I$ where I denotes the identity transformation on $\mathbb{R}^k$. This implies that α must be an orthogonal transformation and that
$$\sum_{j=0}^{m-1} \alpha^j(a) = 0\ . \qquad \text{(A.1)}$$
It follows that if we set $T(P_n^*) = \mu$, we must have $A(\mu) = \mu$ for any affine transformation for which $P_n^{*A} = P_n^*$. The choice of μ is arbitrary subject only to these constraints. Having chosen such a μ the values of $T(P_n^{*B})$ are defined to be $B(\mu)$ for all other affine transformations B.

The above argument shows the internal consistency relationships which must be placed on T so that $T(P_n^{*A}) = A\big(T(P_n)\big)$ for any affine transformation A. We now consider what one may call the external restrictions.

- **Case 3:** Suppose that P_n^* is such that there does not exist a $P_n'^*$ and an affine transformation A such that $P_n^{*A} = P_n'^*$. In this case the choice of $T(P_n^*)$ is only restricted by the considerations of Case 2 above if that case applies and otherwise not at all.

- **Case 4:** Suppose that P_n^* is such that there exists a $P_n'^*$ and an affine transformation A such that $P_n^* = P_n'^{*A}$. In this case we require $T(P_n^*) = A\big(T(P_n'^*)\big)$.

We now place the following conditions on the sample $\boldsymbol{x}_n$:

Sample condition 1: There do not exist two distinct subsets of $\boldsymbol{x}_n$ each of size at least $k+2$ and an affine transformation A which transforms one subset into the other.

Sample condition 2: If

$$\big|A(\boldsymbol{x}_n) \cap B(\boldsymbol{x}_n)\big| \;\geq\; \lfloor (n+1)/2 \rfloor - 2k$$

for two affine transformations A and B then $A = B$.

Sample condition 3: $k < \lfloor (n-1)/2 \rfloor$.

We now construct a functional T which satisfies (4.2), (4.3) and (4.4). If the sample conditions hold then for any affine transformation $A \neq I$ we have $P_n^A \neq P_n^*$ where P_n^* derives from a subset $\boldsymbol{x}_n^*$ which differs from $\boldsymbol{x}_n$ by at least one and at most $\lfloor (n-1)/2 \rfloor$ points. This follows on noting that at most $k+1$ of the $A(x_i)$ belong to $\boldsymbol{x}_n$ by Sample condition 1. Because of this we can define $T(P_n)$ without reference to the values of $T(P_n^*)$. We set

$$T(P_n) \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \; .$$

If P_n^* satisfies the conditions of Case 3 above we set

$$T(P_n^*) \;=\; \frac{1}{n^*}\sum_{i=1}^{n^*} x_{\pi(i)}$$

where the $x_{\pi(i)}$ are those $n^* \geq \lceil (n+1)/2 \rceil$ points of the sample $\boldsymbol{x}_n$ which also belong to the sample $\boldsymbol{x}_n^*$. Finally we consider Case 4 above. We show that the sample assumptions and the condition $P_n^* = P_n^{\prime *A}$ uniquely determine the affine transformation A. To see this we suppose that there exists a second affine transformation B and a distribution $P_n^{\prime\prime *}$ such that $P_n^* = P_n^{\prime\prime *B}$. Let $x^*_{\pi(1)}, ..., x^*_{\pi(N')}$ denote those points of $\boldsymbol{x}_n^*$ not contained in the sample $\boldsymbol{x}_n$. Because of Sample condition 1 this set contains at least $\lceil (n+1)/2 \rceil - k - 2$ points of the form $A(x_i)$. Similarly it also contains at least $\lceil (n+1)/2 \rceil - k - 2$ points of the form $B(x_i)$. The intersection of these two sets is of size at least $\lfloor (n+1)/2 \rfloor - 2k$ and we may conclude from Sample condition 2 that $A = B$. The representation is therefore unique. Let $x_{\pi(1)}, ..., x_{\pi(m)}$ be those points of $\boldsymbol{x}_n$ which belong to the sample $\boldsymbol{x}_n^{\prime *}$ and for which $A(x_{\pi(1)}), ..., A(x_{\pi(m)})$ belong to the sample $\boldsymbol{x}_n$. It is clear that $m \geq 1$. We define

$$T(P_n^{\prime *}) \;=\; \frac{1}{m}\sum_{i=1}^{m} x_{\pi(i)}$$

and by equivariance

$$T(P_n^{*}) \;=\; \frac{1}{m}\sum_{i=1}^{m} \mathcal{A}\big(x_{\pi(i)}\big) \; .$$

It follows that $T(P_n^*)$ is well defined and in both cases the sums involved come from the sample $\boldsymbol{x}_n$. The functional T is extended to all $P_n{*}^B$ and $P_n'{*}^B$ by affine equivariance. In all cases the definition of $T(P_n^*)$ is as the mean of a subset of $\boldsymbol{x}_n$. From this it is clear that (4.4) is satisfied.

APPENDIX B

We now show that Sample conditions 1 and 2 hold for independent random samples $X_1, ..., X_n$ with probability one. Let $\mathcal{A} = A + a$ and $\mathcal{B} = B + b$ with A and B nonsingular matrices and a and b points in $\mathbb{R}^k$. We suppose that $A \neq B$. On taking differences we see that there exist variables $X_{i_1}, ..., X_{i_{k+1}}$ and $X_{j_1}, ..., X_{j_{k+1}}$ such that

$$A(X_{i_l} - X_{i_{k+1}}) = B(X_{j_l} - X_{j_{k+1}}) \ , \qquad j = 1, ..., k \ .$$

This implies that $B^{-1}A$ and $B^{-1}(b-a)$ are functions of the chosen sample points

$$\begin{aligned} B^{-1}A &= C\big(X_{i_1}, ..., X_{i_{k+1}},\, X_{j_1}, ..., X_{j_{k+1}}\big) \ , \\ B^{-1}(b-a) &= c\big(X_{i_1}, ..., X_{i_{k+1}},\, X_{j_1}, ..., X_{j_{k+1}}\big) \ . \end{aligned} \tag{B.1}$$

For n sufficiently large there exist four further sample points X_i, $i = 1, ..., 4$ which are not contained in $\{X_{i_1}, ..., X_{i_{k+1}},\, X_{j_1}, ..., X_{j_{k+1}}\}$ and for which

$$A(X_1) + a \ = \ B(X_2) + b \ , \qquad A(X_3) + a \ = \ B(X_4) + b \ .$$

This implies

$$B^{-1}A(X_3 - X_1) \ = \ X_4 - X_2 \ . \tag{B.2}$$

However as the X_i, $i = 1, ..., 4$, are independent of $X_{i_1}, ..., X_{i_{k+1}}, X_{j_1}, ..., X_{j_{k+1}}$ it follows from (B.1) that (B.2) holds with probability zero. From this we conclude that $A = B$. Similarly we can show that $a = b$ and hence $\mathcal{A} = \mathcal{B}$.

ACKNOWLEDGMENTS

Research supported in part by Sonderforschungsbereich 475, University of Dortmund.

REFERENCES

[1] Croux, C. and Rousseeuw, P.J. (1992). A class of high-breakdown estimators based on subranges, *Commun. Statist. – Theory Meth.*, **21**, 1935–1951.

[2] Davies, P.L. (1993). Aspects of robust linear regression, *Ann. Statist.*, **21**, 1843–1899.

[3] Davies, P.L. and Gather, U. (1993). The identification of multiple outliers (with discussion and rejoinder), *J. Amer. Statist. Assoc.*, **88**, 782–801.

[4] Davies, P.L. and Gather, U. (2005). Breakdown and groups (with discussion and rejoinder), *Ann. Statist.*, **33**, 977–1035.

[5] Davies, P.L. and Gather, U. (2006). Addendum to the discussion of breakdown and groups, *Ann. Statist.*, **34**, 1577–1579.

[6] Davies, P.L. and Kovac, A. (2004). Densities, spectral densities and modality, *Ann. Statist.*, **32**, 1093–1136.

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

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

[10] Ellis, S.P. and Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression, *J. Amer. Statist. Assoc.*, **87**, 143–148.

[11] Gordaliza, A. (1991). On the breakdown point of multivariate location estimators based on trimming procedures, *Statist. Probab. Letters*, **11**, 387–394.

[12] Hampel, F.R. (1968). *Contributions to the theory of robust estimation*, Ph. D. thesis, Dept. Statistics, Univ. California, Berkeley.

[13] Hampel, F.R. (1971). A general qualitative definition of robustness, *Ann. Math. Statist.*, **42**, 1887–1896.

[14] Hampel, F.R. (1975). Beyond location parameters: robust concepts and methods (with discussion), *Proceedings of the 40th Session of the ISI*, Vol. 46, Book 1, 375–391.

[15] He, X. and Fung, W.K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis, *J. Multivariate Anal.*, **72**, 151–162.

[16] Huber, P.J. (1981). *Robust Statistics*, Wiley, New York.

[17] Hubert, M. (1997). The breakdown value of the L_1 estimator in contingency tables, *Statist. Probab. Letters*, **33**, 419–425.

[18] Lopuhaä, H.P. (1992). Highly efficient estimators of multivariate location with high breakdown point, *Ann. Statist.*, **20**, 398–413.

[19] 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.

[20] Müller, C.H. and Uhlig, S. (2001). Estimation of variance components with high breakdown point and high efficiency, *Biometrika*, **88**, 353–366.

[21] Niinimaa, A.; Oja, H. and Tableman, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version, *Statist. Probab. Letters*, **10**, 325–328.

[22] Rousseeuw, P.J. (1984). Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871–880.

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

[24] Terbeck, W. and Davies, P.L. (1998). Interactions and outliers in the two-way analysis of variance, *Ann. Statist.*, **26**, 1279–1305.