

# Maximum Deviation Curves for Location Estimators

Christophe Croux \*

Université Libre de Bruxelles

*Abstract:* The maximum deviation curve of an estimator describes how an estimate can change (in the worst case) when you replace  $m$  out of  $n$  “good” observations to arbitrary positions. This function will be computed for some robust univariate location estimators. A lower bound for this curve is derived, and it is shown that this bound can be attained. Trimmed means will always be close to this lower bound. When more than one third of the observations is contaminated, the median also gets to the lower bound. Finally, it is shown that a high breakdown point leads to a relatively large maximum deviation in the presence of small amounts of contaminants. The maximum deviation curve approach is based on the finite sample behavior of the estimators and makes no distributional assumptions.

*Key words:* Bias curve, Breakdown point, Location estimator, Maximum Deviation, Robustness, Sensitivity.

*AMS subject classifications:* 62F35, 62G35

---

\*C.E.M.E. and Institut de Statistique, Université Libre de Bruxelles, CP 139, Av. F.D. Roosevelt 50, B-1050 Bruxelles, Belgium

# 1 Introduction

In robust statistics one often tries to estimate a parameter from observations following a distribution belonging to a certain neighborhood of a parametric family. Another feature of robust statistics is to detect structure in the data, even before a parametric model is specified. In that case, one assumes that the majority of the observations belongs to a certain “structure”. A minority of the points however may belong to other structures and will be called contaminants. Estimators can help us to identify and summarize the main structure of the data. Therefore it is important that these estimators are quite stable when a certain proportion of the data is contaminated. The aim of this article is to investigate the maximum deviation of an estimator under contamination. For reasons of simplicity, only univariate location estimators are considered. It is important to notice that no parametric assumptions will be made, since we focus in this paper on the behavior of robust estimators in the context of an exploratory data-analysis.

Consider uniquely defined location estimators  $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$ , symmetric in their arguments and satisfying the following affine equivariance property:

$$T_n(aX + b) = aT_n(X) + b \quad \text{for all real numbers } a \text{ and } b \quad (1.1)$$

and for every data set  $X = \{x_1, \dots, x_n\}$ .

A well known measure of robustness is the sensitivity curve ( $SC$ ) of an estimator defined as

$$SC(x, X, T_n) = (n + 1)\{T_{n+1}(X \cup \{x\}) - T_n(X)\}.$$

The sensitivity curve measures the influence of one single observation on the estimator  $T_n$ . It can be seen as a finite sample version of the influence function (Hampel et al. 1986, page 43). The drawbacks of the sensitivity curve are that it only allows for one outlier and that it depends on the configuration  $X$ . Sometimes  $X$  is taken to be a stylized sample (Andrews et al. 1972, page 96) from a certain distribution, but this introduces a dependency on the specified distributional form.

Another measure of robustness is the breakdown point (Donoho and Huber 1983) of an estimator. It gives the minimal fraction of observations you have to replace before the estimator tends to infinity and it is defined as

$$\varepsilon^*(X, T_n) = \inf\left\{\frac{m}{n}; \sup_{X' \in N_m(X)} |T_n(X) - T_n(X')| = \infty\right\}$$

where  $N_m(X)$  is the collection of all data sets obtained by replacing  $m$  observations from  $X$  to arbitrary places. For most (but not all) location estimators the breakdown point does not depend on  $X$ . Therefore the following definition of the breakdown point of an estimator  $T_n$  will be used in this paper:

$$\varepsilon^*(T_n) = \inf_X \varepsilon^*(X, T_n).$$

The breakdown point tells us when an estimator becomes completely unreliable, i.e. tends to infinity due to contaminating points. It gives a one-number summary of the robustness of the estimator. However, we are looking for more information: what can happen in the worst case when the number of contaminants is between 1 and  $n\varepsilon^*(T_n)$ ? The *maximum deviation curve* (MDC) addresses to this question. It is related to the concept of a bias curve (Hampel et al. 1986, Martin et al. 1989). The bias curve for an estimator  $T$  at a distribution  $F$  describes the worst possible deviation when  $\varepsilon$  percent of the observations are outliers, while  $1 - \varepsilon$  percent of the observations come from the model distribution  $F$ . The maximum deviation curve is essentially different, since it requires no distributional assumptions and describes only the finite sample behavior of the estimator. Denote the maximum deviation of the estimator  $T_n$  at a given data set  $X$  when replacing  $m$  observations to arbitrary positions, by

$$b(m, X, T_n) = \sup\{|T_n(X) - T_n(X')|; X' \in N_m(X)\}. \quad (1.2)$$

Since we want to know something about the behavior of our estimator on all possible data sets  $X$ , we remove the dependency on  $X$ , and obtain the following definition of the maximum deviation curve.

**Definition:**

$$\text{MDC}(m, T_n) = \sup\{b(m, X, T_n); -\frac{1}{2} \leq x_i \leq \frac{1}{2} \text{ for all } x_i \in X\}. \quad (1.3)$$

A condition has been set on the range of the good observations to avoid that the maximum deviation would always be infinite: all the “good” observations have to lie in the interval  $[-1/2, 1/2]$ . This restriction is not really strong due to the equivariance property (1.1). Indeed, suppose for a moment that the  $n$  good observations lie in an interval of length  $R$  and that at most  $m$  out of the  $n$  good observations will become contaminated, then the difference between the estimate based on the clean and the contaminated data can be at most  $R \times \text{MDC}_n(m, T)$ . In order to compute the maximum deviation curve we have to choose both the good observations and the contaminants in such a way that the deviation

$|T_n(X) - T_n(X')|$  becomes maximal. This is a very pessimistic viewpoint, yielding a measure of “robustness under all circumstances” of an estimator.

The maximum deviation curve is finite if  $0 \leq m < n\varepsilon_n^*(T)$ , since

$$\varepsilon^*(T_n) = \inf\left\{\frac{m}{n}; \text{MDC}(m, T_n) = \infty\right\}. \quad (1.4)$$

This means that we can read off the value of the breakdown point from the MDC.

Consider now a sequence of estimators  $\{T_n; n \geq 1\}$ , which we will often abbreviate as “an estimator  $T$ ”. An asymptotic version of the maximum deviation curve of this estimator  $T$  is defined as

$$\text{AMDC}(\varepsilon, T) = \limsup_{n \rightarrow \infty} \text{MDC}(\lfloor \varepsilon n \rfloor, T_n) \quad (1.5)$$

where  $\varepsilon < 1/2$  and  $\lfloor x \rfloor$  denotes the integer part of  $x$ . While the asymptotic breakdown point of an estimator is usually defined as

$$\varepsilon^*(T) = \liminf_n \varepsilon^*(T_n), \quad (1.6)$$

one can also consider the “asymptotic maximum deviation curve” breakdown point

$$\varepsilon_{\text{AMDC}}^*(T) = \inf\{\varepsilon; \text{AMDC}(\varepsilon, T) = \infty\}. \quad (1.7)$$

For most estimators the two breakdown points (1.6) and (1.7) are equal, but in Section 4 a counterexample is given.

In Section 2 we derive a lower bound for the maximum deviation curve, compute the MDC for L-estimators and show that these can attain, with properly chosen scores, the lower bound. The trimmed mean comes very close to this lower bound. For contamination percentages higher than one third, also the median has minimax deviation. It is shown in Section 3 that high breakdown estimators are relatively unstable in the presence of small amounts of contamination. In Section 4 we compute the MDC for some other estimators. A small simulation study and some discussion is presented in the final section.

## 2 A Lower Bound for the Maximum Deviation Curve

A result of Rousseeuw and Leroy (1984, page 185) says that the maximal breakdown point for an affine equivariant location estimator is given by  $\lfloor (n+1)/2 \rfloor / n$ . Due to (1.4) one shall therefore always assume  $1 \leq m < \lfloor (n+1)/2 \rfloor$  when considering  $\text{MDC}(m, T_n)$ .

**Lemma 1.** For every equivariant location estimator  $T_n$  with breakdown point  $\varepsilon^*(T_n)$  we have

$$\text{MDC}(m, T_n) \geq 1 / \left\lceil \frac{n - 2\varepsilon_n^*(T)n + 2}{m} \right\rceil \quad (2.1)$$

where  $\lceil x \rceil$  denotes the smallest integer greater than or equal to  $x$ .

**Proof:** Denote  $d = n\varepsilon^*(T_n) - 1$ . Since  $\text{MDC}(m, T_n)$  is only finite for  $m \leq d$  one may suppose that  $m \leq d$ . The following exact fit property (Rousseeuw and Leroy 1984, page 123) holds:

(EF) If  $n - d$  or more observations of a data set  $Y$  are identical to  $\alpha$ , then  $T_n(Y) = \alpha$ .

Indeed, suppose that  $T_n(Y) \neq \alpha$ , then  $Y'$  defined by  $y'_i = c(y_i - \alpha)$  belongs to  $N_d(\{0, \dots, 0\})$  while  $T_n(Y') = c(T_n(Y) - \alpha) \rightarrow \infty$  if  $c \rightarrow \infty$ . This is in contradiction with  $\text{MDC}_n(d, T) < \infty$ .

Denote now  $X^i$  the data set consisting of  $i$  times  $(-1/2)$  and  $n - i$  times  $1/2$ . Consider the telescope sum

$$T_n(X^d) = \sum_{j=1}^{\lceil (n-2d)/m \rceil} \left( T_n(X^{d+(j-1)m}) - T_n(X^{jm+d}) \right) + T_n(X^{\lceil (n-2d)/m \rceil m + d}). \quad (2.2)$$

Due to (EF)  $T_n(X^d) = 1/2$  and  $T_n(X^{\lceil (n-2d)/m \rceil m + d}) = -1/2$  (since  $\lceil (n-2d)/m \rceil m + d \geq n-d$ ). Furthermore, by definition (1.3) it holds that  $|T_n(X^{jm}) - T_n(X^{(j+1)m})| \leq \text{MDC}(m, T_n)$ . Therefore (2.2) yields

$$\frac{1}{2} \leq \lceil (n-2d)/m \rceil \text{MDC}(m, T_n) - \frac{1}{2}$$

which proves (2.1). □

**Theorem 1.** For every equivariant location estimator  $T_n$  we have

$$\text{MDC}(m, T_n) \geq 1 / \left( \left\lceil \frac{n}{m} \right\rceil - 2 \right). \quad (2.3)$$

**Proof:** The maximum deviation  $\text{MDC}(m, T_n)$  is only finite when  $\varepsilon^*(T_n) \geq m+1$ . The result follows now immediately from Lemma 1. □

In Figure 1 we plotted the lower bound  $1/(\lceil n/m \rceil - 2)$  for a few values of  $n$ . If  $m \geq n/3$  the MDC will be bigger than 1, and if  $n/3 > m \geq n/4$  the MDC is bounded below by  $1/2$ . From Theorem 1 and (1.5) it follows that

$$\text{AMDC}(\varepsilon, T) \geq 1 / \left( \left\lceil \frac{1}{\varepsilon} \right\rceil - 2 \right) \quad \text{for } \varepsilon < \frac{1}{2}. \quad (2.4)$$

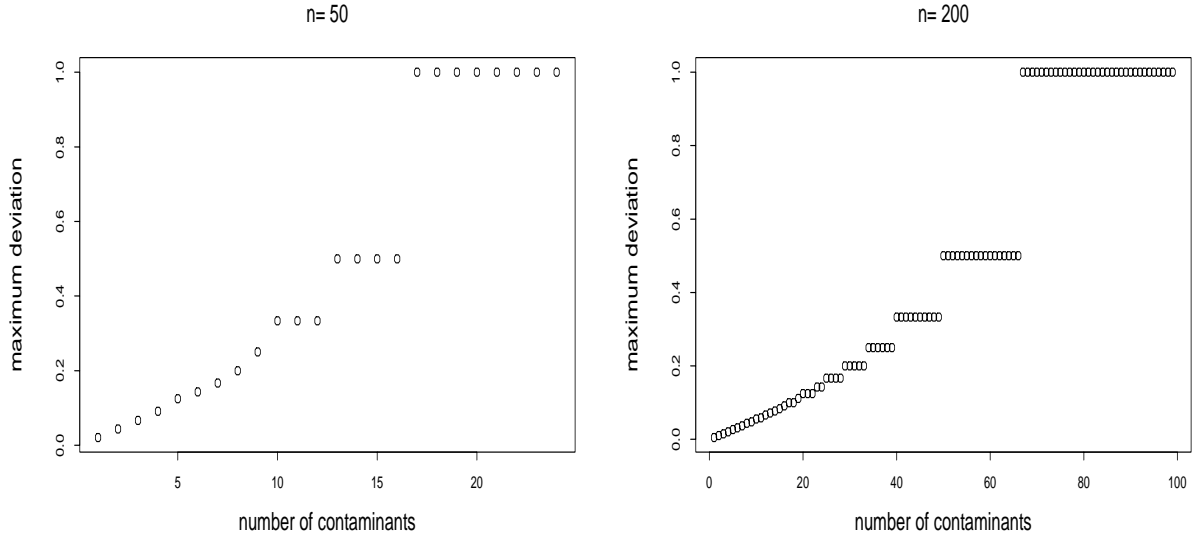


Figure 1: Lower bounds for the maximum deviation curve  $MDC(m, T_n)$  for  $n = 50$  and  $n = 200$

The lower bound increases with the percentage of contamination, but is not strictly increasing.

We will now derive the MDC for the class of L-estimators, which are defined as linear combinations of order statistics:

$$T_n(X) = \sum_{i=1}^n w_i x_{(i)} \quad (2.5)$$

where  $x_{(1)} \leq \dots \leq x_{(n)}$ . In order to obtain the equivariance property (1.1) the weights have to satisfy

$$w_i \geq 0, \quad \sum_{i=1}^n w_i = 1, \quad \text{and} \quad w_i = w_{n-i+1}. \quad (2.6)$$

The tolerance (Hettmansperger 1984, page 18) of an L-estimator is given by  $r = \min\{i; w_i \neq 0\} - 1$  and satisfies  $r \leq \lfloor (n-1)/2 \rfloor$ . It is easy to see that  $r < m$  implies  $MDC(m, T_n) = \infty$ . The next theorem shows that the  $MDC(T_n, m)$  is given by the maximal sum of  $m$  contiguous scores  $w_l, \dots, w_{l+m-1}$ .

**Theorem 2.** *If  $T_n$  is an L-estimator with tolerance  $r \geq m$ , then*

$$\begin{aligned} \text{MDC}(m, T_n) &= \max_{r+1 \leq l \leq n-r-m+1} \sum_{i=l}^{l+m-1} w_i && \text{if } m < n - 2r \\ &= 1 && \text{if } m \geq n - 2r. \end{aligned} \quad (2.7)$$

**Proof:** L-estimators are monotone in the observations, meaning that  $T_n(X) \leq T_n(X')$  if  $x_i \leq x'_i$  for all  $i$ . Therefore, the largest possible difference  $|T_n(X) - T_n(X')|$  for  $X' \in N_m(X)$  is obtained by replacing the  $m$  smallest observations from  $X$  to infinity (or equivalently the  $m$  largest to minus infinity). We will work with that  $X'$ . Then

$$T_n(X') = \sum_{i=r+1}^{n-r} w_i x'_i = \sum_{i=r+1}^{n-r} w_i x_{(i+m)} = \sum_{i=r+1+m}^{n-r+m} w_{i-m} x_{(i)}. \quad (2.8)$$

If  $r + m < n - r$  it follows from (2.8) that

$$T_n(X') - T_n(X) = \sum_{i=n-r+1}^{n-r+m} w_{i-m} x_{(i)} - \sum_{i=r+1}^{r+m} w_i x_{(i)} + \sum_{i=r+m+1}^{n-r} (w_{i-m} - w_i) x_{(i)}. \quad (2.9)$$

Using  $x_{(i)} \leq 1/2$ ,  $x_{(i)} \geq -1/2$  for all  $i$ , and the symmetry of the weights, one obtains for the first two terms in (2.9):

$$\sum_{i=n-r+1}^{n-r+m} w_{i-m} x_{(i)} - \sum_{i=r+1}^{r+m} w_i x_{(i)} \leq \frac{1}{2} \sum_{i=n-r+1}^{n-r+m} w_{i-m} + \frac{1}{2} \sum_{i=r+1}^{r+m} w_i = \sum_{i=r+1}^{r+m} w_i. \quad (2.10)$$

With  $x_{(0)} = -1/2$  the third term in (2.9) equals

$$\sum_{i=r+m+1}^{n-r} (w_{i-m} - w_i) x_{(i)} = \sum_{i=r+m+1}^{n-r} (w_{i-m} - w_i) \left( x_{(0)} + \sum_{l=1}^i (x_{(l)} - x_{(l-1)}) \right) \quad (2.11)$$

Changing the order of the double sum and using  $\sum_{i=r+m+1}^{n-r} (w_{i-m} - w_i) = 0$  yields for (2.11)

$$\sum_{l=1}^{n-r} \left( \sum_{i=\max(l, m+r+1)}^{n-r} (w_{i-m} - w_i) \right) (x_{(l)} - x_{(l-1)}).$$

Since every element  $x_{(j)} - x_{(j-1)}$  ( $j = 1, \dots, n-r$ ) is non-negative and their sum equals  $x_{(n-r)} + 1/2$ , an upper bound for the above expression is given by

$$\left( \max_{l=1, \dots, n-r} \sum_{i=\max(l, m+r+1)}^{n-r} (w_{i-m} - w_i) \right) (x_{(n-r)} + 1/2) \quad (2.12)$$

Since

$$\begin{aligned} \max_{l=1, \dots, n-r} \sum_{i=\max(l, m+r+1)}^{n-r} (w_{i-m} - w_i) &= \max_{l=m+r+1, \dots, n-r} \sum_{i=l}^{n-r} (w_{i-m} - w_i) = \\ \max_{l=m+r+1, \dots, n-r} \sum_{i=l-m}^{l-1} w_i - \sum_{i=n-r-m+1}^{n-r} w_i &= \max_{l=m+r+1, \dots, n-r} \sum_{i=l-m}^{l-1} w_i - \sum_{i=r+1}^{r+m} w_i \end{aligned}$$

and using (2.9), (2.10), and (2.12) one obtains

$$\begin{aligned} T_n(X') - T_n(X) &\leq \left( \sum_{i=r+1}^{r+m} w_i \right) (1/2 - x_{(n-r)}) + \left( \max_{l=m+r+1, \dots, n-r} \sum_{i=l-m}^{l-1} w_i \right) (x_{(n-r)} + 1/2) \\ &\leq \max_{l=m+r+1, \dots, n-r} \sum_{i=l-m}^{l-1} w_i = \max_{l=r+1, \dots, n-r-m+1} \sum_{i=l}^{l+m-1} w_i. \end{aligned}$$

Note that the above inequality becomes an equality if  $x_{(1)} = \dots = x_{(l+m-1)} = -1/2$  and  $x_{(l+m)}, \dots, x_{(n)} = 1/2$  for the choice of  $l$  yielding the maximum in (2.7). This ends the proof for  $m + r \leq n - r$ .

Suppose now that  $m + r \geq n - r$ , then things become easier since

$$T_n(X') - T_n(X) \leq \sum_{i=r+1}^{n-r} w_i (x_{(i+m)} - x_{(i)}) \leq \sum_{i=r+1}^{n-r} w_i = 1$$

with equality if  $x_{(1)} = \dots = x_{(r+1)} = \dots = x_{(n-r)} = -1/2$  and  $x_{(m+r+1)} = \dots = x_{(n-r+m)} = \dots = x_{(n)} = 1/2$ .  $\square$

**Corollary 1:** Consider the  $r/n$ -trimmed mean, given by

$$T_n^r = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_{(i)}.$$

Since all weights  $w_i$  ( $r + 1 \leq i \leq n - r$ ) are equal to  $1/(n - 2r)$ , (2.7) yields

$$\text{MDC}(m, T_n^r) = \min\left(1, \frac{m}{n - 2r}\right) \quad \text{for } m \leq r.$$

The estimator  $T_n^r$  has minimax deviation if  $r \geq m \geq n/3$ , in which case  $\text{MDC}(m, T_n^r) = 1 = 1/(\lceil n/m \rceil - 2)$ . The lower bound can also be reached when  $r$  is a divisor of  $n$  and  $m = r$ , yielding  $\text{MDC}(r, T_n^r) = r/(n - 2r) = 1/(\lceil n/r \rceil - 2)$ .

The asymptotic maximum deviation curve of the  $\alpha$ -trimmed mean, which is the estimator corresponding to the sequence  $\{T_n^{\lfloor \alpha n \rfloor}; n \geq 2\}$ , is given by

$$\text{AMDC}(\varepsilon, T^\alpha) = \min\left(1, \frac{\varepsilon}{1 - 2\alpha}\right) \quad \text{for } \varepsilon \leq \alpha < \frac{1}{2}.$$

Figure 2 pictures  $\text{AMDC}(\varepsilon, T^\varepsilon)$  together with the lower bound. We see that these two curves are close, especially when  $\varepsilon \leq 1/4$ . The function  $\text{AMDC}(\varepsilon, T^\alpha)$  for  $\alpha = 1/4$  is also plotted in Figure 2. This curve is linear in the percentage of contamination and has a low maximum deviation for  $\varepsilon < 1/4$  (the more realistic range of contamination percentages).



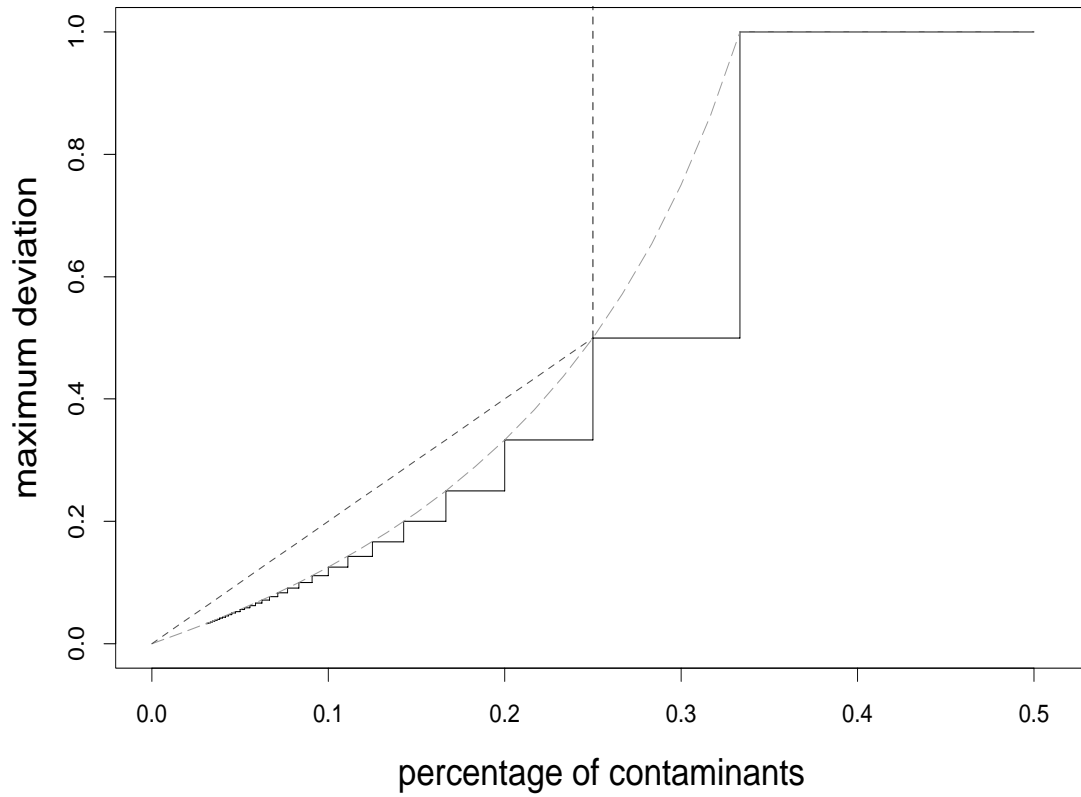


Figure 2: Asymptotic maximum deviation curve for the 25%-trimmed mean (dashed line) together with the lower bound (solid line) and  $AMDC(\varepsilon, T^\varepsilon)$ , where  $T^\varepsilon$  is the  $\varepsilon\%$ -trimmed mean (long dashed line).

**Corollary 2:** Since the median is equal to a  $r/n$ -trimmed mean with  $r = \lfloor (n-1)/2 \rfloor$ , Corollary 1 yields  $\text{MDC}(m, \text{med}_n) = 1$  for  $m \leq r$ . (Except for  $m = 1$  and  $n$  even, then  $\text{MDC}(1, \text{med}_n) = 1/2$ .) It follows that the median has minimal max-deviation for  $m \geq n/3$ .

**Corollary 3:** Define the midpoint of the  $r$ -trimmed data as

$$\text{QM}_n^r = \frac{x_{(r+1)} + x_{(n-r)}}{2} \text{ for } r \leq \lfloor (n-1)/2 \rfloor. \quad (2.13)$$

This estimator was called a *quasi-median* by Mosteller (1946). The maximum deviation curve of this estimator, as given by (2.7), equals

$$\begin{aligned} \text{MDC}(m, \text{QM}_n^r) &= 1 && \text{for } n - 2r \leq m \leq r \\ &= 1/2 && \text{for } 1 \leq m \leq \min(n - 2r - 1, r). \end{aligned} \quad (2.14)$$

From (2.14) we obtain the asymptotic version of the maximum deviation curve for  $\text{QM}^\alpha$

$$\begin{aligned} \text{AMDC}(\varepsilon, \text{QM}^\alpha) &= 1 && \text{for } 1 - 2\alpha \leq \varepsilon \leq \alpha \\ &= \frac{1}{2} && \text{for } 0 < \varepsilon < 1 - 2\alpha \leq \alpha \text{ and } \varepsilon \leq \alpha < 1 - 2\alpha \end{aligned}$$

yielding optimality for  $1/3 \leq \varepsilon \leq \alpha < 1/2$ ,  $1/4 \leq \varepsilon \leq \alpha \leq 1/3$ .

One can see from the above corollaries that there is no unique estimator with minimal  $\text{MDC}(m, T_n)$ . It is possible to define a weighting scheme which defines an L-estimator  $\text{WT}_n^r$  such that  $\text{MDC}(r, \text{WT}_n^r)$  is always optimal. For a fixed  $n$  and  $r \leq \lfloor (n-1)/2 \rfloor$ :

1. Set  $w_i := 0$  for all  $i$  and set  $j := \lfloor \lfloor n/2 \rfloor / r \rfloor$
2. If  $j > 1$  set  $w_{r+1} = w_{2r+1} = \dots = w_{(j-1)r+1} = w_{n-r} = w_{n-2r} = \dots = w_{n-(j-1)r}$  equal to  $\alpha := 1/(\lceil n/r \rceil - 2)$
3. Set  $w_{jr+1} = w_{n-jr} = \beta := (1 - 2\alpha(j-1))/2$ .

It is easy to verify that  $\text{WT}_n^r$  with weights defined as above obeys (2.6) and has tolerance equal to  $r$ .

**Theorem 3.** For a given amount of contamination  $r$ , the maximum deviation curve of  $\text{WT}_n^r$  reaches the lower bound:

$$\text{MDC}(r, \text{WT}_n^r) = 1 / \left( \left\lceil \frac{n}{r} \right\rceil - 2 \right). \quad (2.15)$$

**Proof:** The case  $r \geq n/3$ , and therefore  $r \geq n - 2r$  is, using (2.7), trivial. Therefore, take  $r < n/3$ . Since  $\text{MDC}(r, \text{WT}_n^r) = \max_{r+1 \leq l \leq n-2r+1} \sum_{i=l}^{l+r-1} w_i$ . one has to maximize the sum of  $r$  contiguous scores. If  $|(n-jr) - (jr+1)| \geq r$  then, due to the construction of our estimator, only one term in such a sum will be different from zero, yielding  $\text{MDC}(r, \text{WT}_n^r) = \max(\alpha, \beta)$ . On the other hand, if  $|(n-jr) - (jr+1)| < r$ , we obtain  $\text{MDC}(r, \text{WT}_n^r) = \max(\alpha, 2\beta)$ .

Some easy calculations show that

$$\beta \leq \alpha \Leftrightarrow \lfloor [n/2] / r \rfloor \geq \lceil n/r \rceil / 2 - 1 \quad (2.16)$$

and

$$2\beta \leq \alpha \Leftrightarrow \lfloor [n/2] / r \rfloor \geq \lceil n/r \rceil / 2 - 1/2. \quad (2.17)$$

Now it is not hard to verify that the RHS of (2.16) is always valid and that the RHS of (2.17) holds under the assumption that  $n - 2jr - 1 < r$ . Therefore  $\text{MDC}(r, \text{WT}_n^r) = \alpha$  which ends the proof.  $\square$

Note that we only have optimality for  $m = r$ , not for  $m < r$ . The estimator  $\text{WT}_n^r$  is thus only “pointwise”, but not uniformly optimal. If  $r = 1$ ,  $\text{WT}_n^r$  equals the  $1/n$ -trimmed mean. If  $r > \lfloor n/2 \rfloor / 2$ , we obtain  $\text{WT}_n^r = \text{QM}_n^r$ . If  $\lfloor n/2 \rfloor / 3 < r \leq \lfloor n/2 \rfloor / 2$  we have

$$\text{WT}_n^r = \frac{2x_{(r+1)} + x_{(2r+1)} + x_{(n-2r)} + 2x_{(n-r)}}{6}$$

or

$$\text{WT}_n^r = \frac{x_{(r+1)} + x_{(2r+1)} + x_{(n-2r)} + x_{(n-r)}}{4}$$

(the first type for roughly  $1/5 \leq r/n < 1/4$ , and the second type for  $1/6 \leq r/n < 1/5$ ).

The practical importance of the estimator  $\text{WT}_n^r$  is limited, since it is only optimal for a fixed amount of contamination  $r$ . It is quite unrealistic to assume that the number of contaminants is known in advance. It is however interesting to know that the lower bound for  $\text{MDC}(m, T_n)$  given by Theorem 1, can be attained for each value of  $m$ .

Highly stable estimators are given by the median for heavy contamination ( $> 33\%$ ) and the  $\text{QM}^{1/3}$  for contamination between 25% and 33%. Corollaries 2 and 3 showed that their MDC equals the lower bound in the indicated range. From Figure 2, we can see that the MDC curve of the 25%-trimmed mean is very close to the lower bound for realistic amounts of contamination ( $< 25\%$ ). In this sense, one can consider the 25%-trimmed mean as a “nearly” optimal estimator and recommend it instead of the more artificial estimators  $\text{WT}_n^r$ .

### 3 Breakdown Point versus Maximum Deviation

It was already mentioned in the literature that a 50% breakdown point leads inevitably to a high instability (Rousseeuw 1994). Here it is shown that the results obtained in Section 2 are consistent with the earlier findings that there is a conflict between high breakdown and local stability of an estimator.

Suppose that an estimator  $T_n$  has the maximal breakdown point  $\varepsilon_n^*(T) = \lfloor (n+1)/2 \rfloor$ . It results from Lemma 1 that  $\text{MDC}(m, T_n) \geq 1$  with the exception of  $\text{MDC}(1, T_n) \geq 1/2$ , for  $m = 1$  and  $n$  even. Comparing this lower bound with the MDC of the median obtained in Corollary 2, shows that the MDC of the median is minimal within the class of maximal breakdown estimators.

Nevertheless, replacing only one observation can change the median quite drastically. A measure of the maximal deviation of an estimator  $T$  in the presence of only one contaminant is given by

$$\gamma(T) = \limsup_n n \text{MDC}(1, T_n).$$

(Note that the maximum deviation is standardized by the fraction of contaminants  $1/n$ .) Since the  $1/n$ -trimmed mean minimizes  $\text{MDC}(1, T_n)$ , it will also yield the minimal  $\gamma(T)$ , namely  $\lim_{n \rightarrow \infty} n/(n-2) = 1$ . From (2.1) a lower bound for this measure of sensitivity to small amounts of contamination is obtained:

$$\gamma(T) \geq \frac{1}{1 - 2\varepsilon^*(T)} \tag{3.1}$$

if  $\varepsilon^*(T) > 0$ . Note that the  $\alpha$ -trimmed mean has the lowest possible  $\gamma(T)$  within the class of estimators with an asymptotic breakdown point of  $\varepsilon^* = \alpha$ , since  $\gamma(T^\alpha) = 1/(1 - 2\alpha)$ . The higher the breakdown point, the higher (in general)  $\gamma$ . For  $\varepsilon^*(T) = \frac{1}{2}$ , we even obtain an infinite  $\gamma$ . In particular,  $\text{AMDC}(\varepsilon, T)$  is discontinuous at  $\varepsilon = 0$  if  $T$  has the maximal breakdown point. This unpleasant property can only be avoided when one gives up the 50% breakdown point. For estimators with a strictly positive (but not maximal) breakdown point, it stays possible to combine the high breakdown property with a finite (but possibly high)  $\gamma$ .

This definition of  $\gamma$  is similar to the definition of the gross-error sensitivity (Hampel et al. 1986, page 87). There it was proved that the median has the lowest possible gross-error sensitivity when the data are distributed according to a symmetric distribution. Our ap-

proach shows however that the median can, at certain configurations, become very sensitive to small amounts of contamination (which corresponds with the intuition).

*Remark:* Defining alternatively  $\gamma(T)$  as  $\limsup_{\varepsilon \downarrow 0} \text{AMDC}(\varepsilon, T)/\varepsilon$  yields the same inequality (3.1).

## 4 Examples

It can be quite hard to compute the maximum deviation curve of an estimator, even in the univariate location model. One often has to limit himself deriving only a lower bound for the MDC. The following examples show how the maximum deviation curve can be used to investigate the behavior of robust estimators.

*Example 1:* Compute  $d_i = |x_i - \text{med}_j x_j|$  as a measure of outlyingness of an observation. The estimator  $DT_n^r$  is defined as the average of all observations remaining after deletion of the  $r$  observations with largest  $d_i$ . This estimator resembles the  $r/n$  trimmed mean, but differs from it since it can attain the same breakdown point while discarding only half as much observations.

**Proposition 1.** *The maximum deviation curve of  $DT_n^r$ , with  $r \leq \lfloor (n-1)/2 \rfloor$  satisfies*

$$\text{MDC}(m, DT_n^r) \geq \frac{2m + \min(r, \lfloor \frac{n}{2} \rfloor + 1 - m)}{n - r} \quad \text{for } m \leq r \quad (4.1)$$

**Proof:** Consider the configuration  $X$  consisting of  $\lfloor (n-1)/2 \rfloor + m$  times  $-1/2$  and  $\lfloor n/2 \rfloor + 1 - m$  times  $1/2$ . Replace now  $m$  points from  $-1/2$  to  $3/2 - \delta$ , with  $\delta$  small and positive in order to obtain  $X'$ . Now it is straightforward to check that  $T_n(X') - T_n(X)$  equals the right hand side in (4.1), when  $\delta$  tends to zero.  $\square$

For the sequence of estimators  $DT_n^{\lfloor \alpha n \rfloor}(X)$  we obtain

$$\text{AMDC}(\varepsilon, DT^\alpha) \geq \frac{2\varepsilon + \min(\alpha, \frac{1}{2} - \varepsilon)}{1 - \alpha} \quad \text{for } 0 < \varepsilon < \frac{1}{2}.$$

The sensitivity to small amounts of contaminations equals  $\gamma(DT^\alpha) = \infty$ . This means that a procedure based on robust deletion of outliers, followed by the classical averaging can be more sensitive to contamination than non data-based deletion. A reason for this is that good points can be detected erroneously as outliers, when the contaminated points are not too far from the good observations. This is called the swamping effect.

*Example 2:* In regression analysis the *Least Median of Squares* (LMS) estimator (Rousseeuw 1984) is a popular highly robust estimator. Taking only an intercept term and no slope parameter reduces the regression model to the univariate location model and the LMS to the midpoint of the shortest interval containing half of the observations. A bit more general, define the univariate  $\text{LMS}_n^r$  as the midpoint of the shortest interval containing  $n - r$  observations, with  $r \leq \lfloor (n - 1)/2 \rfloor$ . The originally proposed LMS takes  $r = \lfloor (n - 1)/2 \rfloor$ . To ensure that  $\text{LMS}_n^r$  is uniquely defined we can take the average over the midpoints of the shortest intervals. It's easy to see that  $\varepsilon^*(\text{LMS}_n^r) = r + 1$ . In particular  $\text{LMS}_n := \text{LMS}_n^{\lfloor (n-1)/2 \rfloor}$  has the maximal breakdown point.

**Proposition 2.** *The maximum deviation curve of the estimator  $\text{LMS}_n^r$  is given by*

$$\begin{aligned} \text{MDC}(m, \text{LMS}_n^r) &= \frac{3}{2} && \text{for } n - 2r \leq m \leq r \\ &= 1 && \text{for } 1 \leq m < n - 2r \text{ and } m \leq r \end{aligned}$$

**Proof:** For proving the inequality  $\geq$ , consider the sample  $X$  consisting of  $m + r$  times  $-1/2$  and  $n - r - m$  times  $1/2$ . If  $r + m \geq n - r \leftrightarrow m \geq n - 2r$  then  $\text{LMS}_n^r(X) = -1/2$ , otherwise  $\text{LMS}_n^r(X) = 0$ . Now remove  $m$  points from  $-1/2$  to  $(3 - \delta)/2$  for a small positive  $\delta$ . Then  $\text{LMS}_n^r(X') = (1/2 + (3 - \delta)/2)/2 = 1 - \delta/4$ , which tends to 1 if  $\delta$  tends to 0. Since  $\text{MDC}_n(m, \text{LMS}^r) \geq \inf_{\delta} |\text{LMS}_n^r(X) - \text{LMS}_n^r(X')|$ , the  $\geq$  part is proven.

For the inequality  $\leq$  consider an arbitrary sample  $X$  in the interval  $[-1/2, 1/2]$ , and replace now  $m$  arbitrary points to arbitrary positions to get  $X'$ . First, there is an interval containing at least  $n - r \leq n - m$  observations lying in  $[-1/2, 1/2]$  with length at most 1. Secondly, the shortest interval should have at least  $n - r - m \geq 1$  points in common with  $[-1/2, 1/2]$ . Combining these two facts gives that the shortest interval covering half of the points of  $X'$  lies somewhere between  $-3/2$  and  $3/2$  and has a non-empty intersection with  $[-1/2, 1/2]$ . Therefore  $|\text{LMS}_n^r(X')| \leq 1$  yielding  $|\text{LMS}_n^r(X) - \text{LMS}_n^r(X')| \leq 3/2$ . Now the shortest intervals based on the clean observations and on the contaminated observations have at least  $(n - r - m) - r = n - 2r - m$  points in common. If  $m < n - 2r$  both intervals have a non-empty intersection and a length smaller than 1, yielding  $|\text{LMS}_n^r(X) - \text{LMS}_n^r(X')| \leq 1$ .  $\square$

For the asymptotic version  $\text{LMS}^\alpha$  (corresponding to the sequence  $\{\text{LMS}_n^{\lfloor \alpha n \rfloor}; n \geq 2\}$ ) we obtain

$$\text{AMDC}(\varepsilon, \text{LMS}^\alpha) = \frac{3}{2} \quad \text{for } 1 - 2\alpha \leq \varepsilon \leq \alpha \text{ and } \alpha \geq 1/3$$

$$\begin{aligned}
&= 1 \quad \text{for } \varepsilon < 1 - 2\alpha \leq \alpha \text{ and } \alpha \geq 1/3 \\
&= 1 \quad \text{for } \alpha < 1/3.
\end{aligned}$$

As a consequence the LMS estimator and the median, which both have the same breakdown point, don't have the same maximum deviation curve. Also note that  $\gamma(\text{LMS}^\alpha)$  is infinite for each value of  $\alpha$ .

*Example 3:* This example will illustrate the power of the maximum curve with respect to the breakdown point. Consider the following estimator, based on iterative detection of outliers. Denote  $S_n^0 = \text{med}_i |x_i - \text{med}_j x_j|$  the median absolute deviation scale estimator and  $T_n^0$  the median. We will define  $V^i = \{x_i; |x_i - T_n^0| \leq CS_n^i\}$  where  $S_n^i = \text{range}(V^{i-1})$  for  $i \geq 1$ . When  $V^i = V^{i-1} := V$  for some  $i$  we stop the iteration and compute the average  $T_n$  over the observations in  $V$ . The constant  $C$  is taken larger than 1. It is not difficult to see that the breakdown point (1.4) of this estimator  $T_n$  is the same as that of the initial estimator, the median, namely  $\lfloor (n+1)/2 \rfloor$ .

Replace now  $m$  observations from a data set  $X$  with  $x_{(1)} = -1/2$  and  $x_{(n)} = 1/2$  in order to get a new data set  $X' = \{-1/2, \dots, 1/2, x'_1, \dots, x'_m\}$  with  $x'_i = C^i$ ,  $\text{med}(X') = 0$  and  $V^0 = X \cap X'$ . It is possible to construct such an  $X$  and  $X'$ . Since  $V^i \supseteq (X \cap X') \cup \{x'_1, \dots, x'_i\}$  we have  $V = X'$ . Therefore,  $|T_n(X') - T_n(X)| \geq C^m/n$ . So we may conclude that  $\text{MDC}(m, T_n) \geq C^m/n$ , which yields a fast increasing function in  $m$ . The asymptotic maximum deviation curve satisfies

$$\text{AMDC}(\varepsilon, T) \geq \lim_{n \rightarrow \infty} \frac{C^{\lfloor \varepsilon n \rfloor}}{n} = \infty \text{ for } \varepsilon > 0,$$

yielding  $\varepsilon_{\text{MDC}}^*(T) = 0$ . This seems to be in contradiction with  $\varepsilon^*(T) = 1/2$ . Essentially, this is because the deviation  $\sup_{X'} |T_n(X) - T_n(X')|$  is finite, but not *uniformly* bounded in  $n$ . Therefore the breakdown points defined in (1.6) and (1.7) don't always need to be equal. This example shows that the classical breakdown point can be a very poor measure of the robustness of an estimator (this was also discussed in Croux, 1994). The estimator defined in this example does not deserve to be called robust (although it has a 50% breakdown point), as is witnessed by its large maximum deviation curve and even more by its asymptotic maximum deviation breakdown point.

*Example 4:* A referee asked to compute the maximum deviation curve for the Huber M-estimator of location. This estimator  $H_n$  is defined as the solution of the following equation

in  $t$ :

$$\sum_{i=1}^n \psi_b\left(\frac{x_i - t}{s_n}\right) = 0,$$

where  $\psi_b(u) = \min(\max(b, u), -b)$ . To ensure the affine equivariance property (1.1) an auxiliary scale estimate  $s_n$  is needed. Usually one takes for  $s_n$  the *median absolute deviation* (MAD) about the median:  $s_n = \text{med}_i |x_i - \text{med}_j x_j|$ . In the following proposition a lower bound for the MDC of  $H_n$  is derived. (To avoid technical details, the assumptions  $b > 1$  and  $n$  odd are made.)

**Proposition 3.** *The maximum deviation curve for the above defined Huber M-estimator  $H_n$  with  $b > 1$  and  $n$  odd is given by*

$$\begin{aligned} \text{MDC}(H_n, m) &\geq 1 && \text{for } 1 \leq m \leq h/(1+b) \\ &\geq \frac{1}{2} + \frac{(2b+1)m-1}{2(n-m)} && \text{for } h/(b+1) < m \leq n/2 - (h-1)/2b \\ &\geq 1 + \frac{b(2m-h)}{n-h} && \text{for } n/2 - (h-1)/2b < m \leq \lfloor (n-1)/2 \rfloor \end{aligned}$$

where  $h = \lfloor n/2 \rfloor + 1$ .

**Proof:** Take  $X$  the sample consisting of  $h$  times  $-1/2$  and  $n-h$  times  $1/2$ , yielding  $H_n(X) = -1/2$ . Replace now  $m$  observations from  $-1/2$  to infinity to get  $X'$ . Observe that  $\text{MAD}(X') = 1$  and denote  $\lambda(t) = \sum_{i=1}^n \psi_b(x'_i - t)$ , which is a decreasing function in  $t$ . For  $1/2 \leq t \leq b - 1/2$  we have  $\lambda(t) = mb + (h-m)(-1/2 - t) + (n-h)(1/2 - t)$  and for  $b - 1/2 \leq t \leq b + 1/2$  it yields  $\lambda(t) = mb + (h-m)(-b) + (n-h)(1/2 - t)$ . One can check that  $\lambda(1/2) \leq 0 \Leftrightarrow m \leq h/(1+b)$ ,  $\lambda(b - 1/2) \leq 0 \Leftrightarrow m \leq n/2 - (h-1)/2b$  and  $\lambda(b + 1/2) \leq 0 \Leftrightarrow m \leq n/2$ . If  $\lambda(1/2) \leq 0$  we consider the data set  $X''$  consisting of  $h-m$  times  $-1/2$  and  $n-h+m$  times  $1/2$ , yielding  $\text{MDC}(H_n, m) \geq |H_n(X'') - H_n(X)| = 1$ . For  $\lambda(1/2) > 0 \geq \lambda(b - 1/2)$ , the root of  $\lambda(t) = 0$  is given by  $((2b+1)m-1)/(2(n-m))$ . On the other hand, for  $\lambda(b - 1/2) > 0 \geq \lambda(b + 1/2)$ , the solution of  $\lambda(t) = 0$  is given by  $1/2 + b(2m-h)/(n-h)$ . Noting that  $\lambda(H_n(X')) = 0$  and  $\text{MDC}(H_n, m) \geq |H_n(X') - H_n(X)|$  ends the proof.  $\square$

Since the estimator  $H_n$  has the maximal breakdown point, its MDC is higher than the MDC of the median. We also see that the higher the value of the tuning parameter  $b$ , the higher the maximum deviation. This confirms the well-known fact that the robustness of the Huber-estimator decreases with the value of the tuning parameter  $b$  (but the statistical efficiency at normal distributions will increase).



## References

- Andrews, D.F., P.J. Bickel, P.J., Hampel, F.R., Huber, P.J., Tukey, J.W., and Rogers, W.H. (1972), *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton NJ.
- Croux, C. (1994), “Efficient High-Breakdown M-Estimators of Scale,” *Statistics and Probability Letters*, 19, 371–379.
- Donoho, D.L., and Huber, P.J. (1983), “The Notion of Breakdown Point,” in *A Festschrift for Erich Lehmann*, eds. P. Bickel, K. Doksum, and J.L. Hodges, Jr., Wadsworth, California.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986), *Robust Statistics: the Approach based on Influence Functions*, John Wiley, New York.
- Hettmansperger, T.P. (1984), *Statistical Inference Based on Ranks*, Krieger, Malabar.
- Huber, P.J. (1981), *Robust Statistics*, John Wiley, New York.
- Martin, R.D., V.J. Yohai and R.H. Zamar (1989), Min-max Bias Robust Regression, *Annals of Statistics* 17, 1608–1630.
- Mosteller, F (1946), *Annals of Mathematical Statistics*, 17, 377–408.
- Oosterhoff, J. (1994). “Trimmed Mean or Sample Median?” *Statistics and Probability Letters*, 20, 401-409.
- Riedel, M. (1989), “On the Bias-Robustness in the Location Model I,” *Statistics*, 20, 223-233.
- Rousseeuw, P.J. (1984), Least median of squares regression, *J. Amer. Statist. Assoc.* **79**, 871–880.
- Rousseeuw, P.J. (1994), “Unconventional Features of Positive-Breakdown Estimators,” *Statistics and Probability Letters*, 19, 417–431.
- Rousseeuw, P.J. and A.M. Leroy (1987), *Robust Regression and Outlier Detection*, John Wiley, New York.