April 23, 2003

3.44 Robustness, breakdown points, and 1-dimensional location M-estimators. [Parts of this section are more or less parallel to parts of Section 3.4. One difference is that here, breakdown points are defined in terms of finite samples. Eventually, this section and Section 3.4 will be merged into one or reorganized.]

Let $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ be samples of real numbers. For $j = 1, ..., n$ let $X \ =_j \ Y$ mean that $X_i = Y_i$ except for at most j values of i. More specifically, for $y = (y_1, ..., y_j)$ let $X \ =_{j,y} Y$ mean that for some integers i_r with $1 \leq i_1 < i_2 < ... < i_j \leq n$, $Y_{i_r} = y_r$ for $r = 1, ..., j$ and $Y_i = X_i$ if $i \neq i_r$ for $r = 1, ..., j$. The idea is that in the latter case, X_i are i.i.d. from a nice distribution like a normal and y_r are errors or "bad" data. So the sample X contains $n - j$ good data points and j errors. A robust statistical procedure will be one that doesn't behave too badly if j is not too large compared to n.

"Breakdown point" is one of the main ideas in robustness. Let $T = T(X_1, ..., X_n)$ be a statistic taking values in some locally compact metric space Θ such as a Euclidean space. The closure of a set $A \subset \Theta$ will be denoted $\overline{A}$. If Θ is a Euclidean space then a set $A \subset \Theta$ has compact closure if and only if $\sup\{|x| : \ x \in A\} < \infty$. The *breakdown point* of T at X, or more specifically the *finite-sample breakdown point*, is defined as

$$\varepsilon^*(T, X) \ = \ \varepsilon^*(T; X_1, ..., X_n) = \frac{1}{n} \max\{j : \ \overline{\{T(Y) : \ Y \ =_j \ X\}} \ \text{ is compact}\}.$$

In other words $\varepsilon^*(T, X) = j/n$ for the largest j for which there is some compact set $K \subset \Theta$ such that $T(Y) \in K$ whenever $Y \ =_j \ X$. If $\varepsilon^*(T, X)$ doesn't depend on X, which is often the case, then let $\varepsilon^*(T) \ := \ \varepsilon^*(T, X)$ for all X. If Θ is a Euclidean space $\mathbb{R}^k$, then the compactness condition in the definition of ε^* is equivalent to

$$\sup\{|T(Y)| : \ Y \ =_j X\} \ < \ +\infty.$$

If a fraction of the data less than or equal to the breakdown point is bad (subject to arbitrarily large errors), the statistic doesn't change too much (it remains in a compact set), otherwise it can escape from all compact sets (in a Euclidean space, or by definition in other locally compact spaces, it can go to infinity). There are a number of definitions of breakdown point. The definition of finite-sample breakdown point as above is given in Hampel et al., 1986, p. 98, for a real-valued statistic.

Since j in the definition is an integer, the possible values of the breakdown point for samples of size n are $0, 1/n, 2/n, ..., 1$. A statistic with a breakdown point of 0 is (by definition) not robust. Larger values of the breakdown point indicate more robustness, up to just less than 1/2. Finite-sample breakdown points $\geq 1/2$ are unattainable in some situations, e.g. Theorem 3.44.2 below.

Recall the definition of order statistics: for a sample $X_1, \ldots, X_n$ of real numbers, let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the numbers arranged in order.

Examples. (i) For the sample mean $T = \bar{X} = (X_1 + ... + X_n)/n$, the breakdown point is 0 since for $j = 1$, if we let $y_1 \to \infty$ then $\bar{X} \to \infty$ (for n fixed).
(ii) Let $T = X_{(1)}$, the smallest number in the sample. Then the breakdown point of T is again 0 since for $j = 1$, as $y_1 \to -\infty$ we have $X_{(1)} \to -\infty$. Likewise the maximum $X_{(n)}$ of the sample has breakdown point 0.

So the statistics $\bar{X}$, $X_{(1)}$, $X_{(n)}$ are not robust. Other order statistics have some robustness (for fixed finite n):

Theorem 3.44.1. For sample size n, and each $j = 1, ..., n$, the order statistic $T = X_{(j)}$ has breakdown point $\varepsilon^*(T) = \frac{1}{n}\min(j-1, n-j)$.

Proof. Take any sample $X = (X_1, ..., X_n)$. We have $\inf\{T(Y) : Y =_j X\} = -\infty$ (let $y_1, ..., y_j$ all go to $-\infty$). Likewise $\sup\{T(Y) : Y =_{n-j+1} X\} = +\infty$ (let $y_1, ..., y_{n-j+1} \to +\infty$). So $\varepsilon^*(T, X) \le \frac{1}{n}\min(j-1, n-j)$.

If $Y =_{j-1} X$ then the smallest possible value of $Y_{(j)}$ occurs when $y_i < X_k$ for all i and k and for at least one k such that $X_k = X_{(1)}$, X_k is not replaced, so $Y_{(j)} \ge X_{(1)}$. Similarly, if $Y =_{n-j} X$ the largest possible value of $Y_{(j)}$ satisfies $Y_{(j)} \le X_{(n)}$. So if $r = \min(j-1, n-j)$ and $Y =_r X$, then $X_{(1)} \le Y_{(j)} \le X_{(n)}$ so $Y_{(j)}$ is bounded and $\varepsilon^*(T, X) = \frac{1}{n}\min(j-1, n-j)$ as claimed. Since this is true for an arbitrary X, the theorem is proved. $\square$

If $j = 1$ or n, the breakdown point of $X_{(j)}$ is 0 as noted in the Examples above. If n is odd, so $n = 2k+1$ for an integer k, then the sample median $X_{(k+1)}$ has breakdown point $\frac{1}{2} - \frac{1}{2n} = \frac{k}{n}$. If $n = 2k$ for an integer k, then the two endpoints of the interval of medians, $Y_{(k)}$ and $Y_{(k+1)}$, each have breakdown point $\frac{1}{2} - \frac{1}{n}$. So any median has breakdown point at least $\frac{1}{2} - \frac{1}{n} \to \frac{1}{2}$ as $n \to \infty$. From Theorem 3.44.1, no other order statistic has any larger breakdown point than the median, so $\varepsilon^*(X_{(j)}) < 1/2$ for all j. This is typical behavior for interesting estimators. But, larger breakdown points are possible. If T has values in a compact set, then it trivially has breakdown point 1 by our definition. Or, let $T = \min_j |X_j|$ for observations in $\mathbb{R}^k$. Then one can check that T has breakdown point $1 - \frac{1}{n}$.

For real-valued observations $X_1, \ldots, X_n$, a real-valued statistic $T = T(X_1, ..., X_n)$ will be called *equivariant for location* if for all real θ, and letting $X = (X_1, \ldots, X_n)$ and $X + \theta = (X_1 + \theta, ..., X_n + \theta)$,

$$T(X + \theta) = T(X) + \theta$$

for all n-vectors X of real numbers and all real θ.

For example, the order statistics $X_{(j)}$ and the sample mean $\bar{X}$ are clearly equivariant for location.

Theorem 3.44.2. For any real-valued statistic T equivariant for location, the breakdown point is $< 1/2$ at any $X = (X_1, ..., X_n)$.

Proof. Let the breakdown point of T at X be j/n. Then there is an $M < \infty$ such that

$$|T(Y)| \le M \text{ whenever } Y =_j X. \tag{3.44.3}$$

Let $\theta = 3M$. Now $W = Y + \theta$ for some Y with $Y =_j X$ if and only if $W =_j X + \theta$. Then $T(W) = T(Y) + \theta$. So

$$|T(W) - \theta| \le M \text{ whenever } W =_j X + \theta, \text{ and then } 2M \le T(W) \le 4M. \tag{3.44.4}$$

But if $j \geq n/2$ there is a W with $W =_j X$ and also $W =_j X + \theta$. For such a W, (3.44.3) and (3.44.4) give a contradiction, proving Theorem 3.44.2. □

Now, we'll consider breakdown points of 1-dimensional location M-estimators. Let ψ be a real-valued function of a real variable which is odd (meaning $\psi(-x) \equiv -\psi(x)$), nondecreasing, nonconstant, and bounded. Then $\psi(-t) \leq 0 = \psi(0) \leq \psi(t)$ for all $t \geq 0$ and $\psi(-t) < 0 < \psi(t)$ for some $t > 0$ since ψ is nonconstant. We will have $\psi(t) \to A$ as $t \to +\infty$ for some $A > 0$. Examples of such functions ψ include the derivatives $\rho'(x)$ of wide-sense Huber functions (as defined in §3.4), where such derivatives are defined, with suitable choices where they are not defined, specifically, $\psi(0) = 0$ in all cases, $\psi(x) := \rho'(x+) := \lim_{h \downarrow 0}(\rho(x+h) - \rho(x))/h$ and $\psi(-x) := -\psi(x)$ for $x > 0$. Then for location, the psi function of two variables is defined by $\psi(\theta, x) := \psi(x - \theta)$, which is nonincreasing in θ. Given a sample $(X_1, \ldots, X_n)$, let

$$\theta^* := \theta^*(X_1, ..., X_n) := \sup\left\{\theta : \sum_{i=1}^n \psi(X_i - \theta) > 0\right\}.$$

This is finite since the sum is ≤ 0 for $\theta \geq X_{(n)}$ and also < 0 when $\theta \geq X_{(n)} + t$ for some t such that $\psi(t) > 0$. Analogously, define

$$\theta^{**} := \theta^{**}(X_1, \ldots, X_n) := \inf\left\{\theta : \sum_{i=1}^n \psi(X_i - \theta) < 0\right\},$$

which is also finite since the sum is ≥ 0 for $\theta \leq X_{(1)}$. We have $\theta^* \leq \theta^{**}$ because of the monotonicity of ψ. Then a statistic $T_n = T_n(X_1, \ldots, X_n)$ will be an M-estimator of extended ψ type if and only if $\theta^* \leq T_n \leq \theta^{**}$. In order to have a unique estimator, *the* M-estimator defined by ψ and the sample will be defined, as for the sample median, by

$$\hat{\theta} := \hat{\theta}((x_1, \ldots, x_n)) := \frac{1}{2}(\theta^* + \theta^{**})((x_1, \ldots, x_n)).$$

It will shown that such estimators have the same (finite sample) breakdown points as the median, converging to 1/2 as $n \to \infty$ and as large as possible. Consider also scale-adjusted M-estimators, where instead of $\sum_{i=1}^n \psi(X_i - \theta)$ we have $\sum_{i=1}^n \psi((X_i - \theta)/S)$ and S is a scale estimator, with nonnegative values. The resulting θ^*, θ^{**}, and $\hat{\theta}$ will be called θ_S^*, θ_S^{**}, and $\hat{\theta}_S := (\theta^* + \theta^{**})/2$ respectively. If $S = 0$, then by definition set

$$\psi((X_i - \theta)/S) := \begin{cases} A, & X_i > \theta \\ 0, & X_i = \theta \\ -A, & X_i < \theta. \end{cases}$$

It's easily seen that if $S = 0$ then the M-estimator $\hat{\theta}_S$ based on the above definitions is exactly the median.

To get a particular choice of S, let M be the median of the sample, defined as $X_{(k+1)}$ if $n = 2k+1$ is odd, and $(X_{(k)} + X_{(k+1)})/2$ if $n = 2k$ is even. Let MAD denote the median absolute deviation, namely the median of $|X_i - M|$, and $S = \mathrm{MAD}/.6745$, where the constant 0.6745 is (to the given accuracy) the median of $|Z|$ for a standard normal variable Z, and thus, S estimates the standard deviation σ for normally distributed data.

The following fact and proof are adapted from Huber (1981), pp. 52-53.

Theorem 3.44.5. Let ψ be a function from $\mathbb{R}$ into $\mathbb{R}$, which is odd, nondecreasing, nonconstant, and bounded. Then the M-estimator $\hat{\theta}$ defined by ψ has breakdown point $\frac{1}{2} - \frac{1}{n}$ if n is even and $\frac{1}{2} - \frac{1}{2n}$ if n is odd. The same holds for the scale-adjusted M-estimator $\hat{\theta}_S$ where we consider $\psi((X_i - \theta)/S)$ for the S just defined.

Proof. As $t \to \infty$ we have $\psi(t)\uparrow A$. For $0 < \varepsilon < 1$ there is a $\kappa < \infty$ such that $\psi(\kappa) \geq (1-\varepsilon)A$. Then $\sum_{i=1}^n \psi(X_{(i)} - \theta) < 0$ if $X_{(i)} - \theta < -\kappa$ for j values of i where $-j(1-\varepsilon)A + (n-j)A < 0$, or equivalently $j > n/(2-\varepsilon)$. Now $\theta > X_{(i)} + \kappa$ for at least j values of i is equivalent to $\theta > X_{(j)} + \kappa$. So we have $\theta^{**} \leq X_{(j)} + \kappa$ where j is the smallest integer $> n/(2-\varepsilon)$.

Now if $Y_i = X_i$ for at least j values of $i = 1, ..., n$, we have $Y_{(j)} \leq X_{(n)}$, so $\theta^{**}(Y_1, ..., Y_n) \leq X_{(n)} + \kappa$ and θ^{**} remains bounded above under the given conditions. Symmetrically, θ^* stays bounded below. It follows that $\hat{\theta}$ stays bounded, so the breakdown point of $\hat{\theta}$ is at least $1 - j/n$.

If $i > n/2$, then for some $\varepsilon > 0$, $i > n/(2-\varepsilon)$, so we can take j as the smallest integer greater than $n/2$. Then for $n = 2m$ even, or for $n = 2m+1$ odd, we have $j = m+1$. It follows that the breakdown point of $\hat{\theta}$ is at least as large as stated for each sample size.

$\hat{\theta}$ is a location equivariant estimator, so its breakdown point is less than 1/2 by Theorem 3.44.2. So the breakdown point is no larger than for the median and is the same as for the median, proving the first statement in the theorem.

Next, consider the scale-adjusted case and first, the breakdown point of S. If j is again the smallest integer $> n/2$, and $Y =_{n-j} X$, so that $Y_i = X_i$ for at least j values of i, then $X_{(1)} \leq Y_i \leq X_{(n)}$ for at least j values of i. Thus as noted above

$$Y_{(j)} \leq X_{(n)} \tag{3.44.6}$$

and if M_Y is the median of $Y_1, ..., Y_n$, then $X_{(1)} \leq M_Y \leq X_{(n)}$. Also, $|Y_i - M_Y| \leq X_{(n)} - X_{(1)}$ for at least j values of i, so MAD_Y, the median of $|Y_i - M_Y|$, satisfies $\mathrm{MAD}_Y \leq X_{(n)} - X_{(1)}$ and

$$S_Y \ := \ S(Y_1, ..., Y_n) \leq K_X \ := \ (X_{(n)} - X_{(1)})/0.6745. \tag{3.44.7}$$

On the other hand if $Y =_k X$ for $k \geq n/2$ we can have M_Y unbounded and also MAD_Y unbounded. So the MAD and S have the same breakdown point as the median itself.

It's possible that $S = 0$ if there are enough tied observations. As noted above, the M-estimator $\hat{\theta}_S$ equals the sample median in that case.

Returning to the case where j is the smallest integer $> n/2$ and $Y =_{n-j} X$, take $\varepsilon > 0$ small enough so that $j > n/(2-\varepsilon)$ and choose κ accordingly. Then we will have

$$\sum_{i=1}^n \psi((Y_i - \theta)/S_Y) < 0 \tag{3.44.8}$$

if $Y_i - \theta < -\kappa S_Y$ for at least j values of i, in other words if $Y_{(j)} - \theta < -\kappa S_Y$ or $\theta > Y_{(j)} + \kappa S_Y$. This does hold when $S_Y = 0$: then if $Y_i - \theta < 0$ for at least j values of i, we have by the definitions $\sum_{i=1}^n \psi((Y_i - \theta)/S_Y) \leq -jA + (n-j)A = (n-2j)A < 0$. Now (3.44.8) holds if $\theta > X_{(n)} + \kappa K_X$ by (1) and (2). So

$$\theta^{**}(Y_1, ..., Y_n) \leq X_{(n)} + \kappa K_X.$$

Symmetrically, we have

$$\theta^{*}(Y_1, ..., Y_n) \geq X_{(1)} - \kappa K_X.$$

So $\hat{\theta}(Y_1, ..., Y_n)$ remains bounded for $Y =_{n-j} X$ and the breakdown point of $\hat{\theta}$ is at least as large as for the median. If $S = 0$, this doesn't cause breakdown of the M-estimator because the median of a sample $Y_1, ..., Y_n$ containing more than $n/2$ of the original observations $X_1, ..., X_n$ must be between $X_{(1)}$ and $X_{(n)}$ and so can't become unbounded.

Since $\hat{\theta}_S$ is also location equivariant, its breakdown point is also $< 1/2$ and so equals that of the median, as stated in the Theorem. □

PROBLEMS

1. Let k be a positive integer and $n = 4k+3$. Let $X_1, ..., X_n$ be a sample of real numbers with order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Then the sample median is $X_{(2k+2)}$. Define $X_{(k+1)}$ to be the *lower quartile* and $X_{(3k+3)}$ to be the *upper quartile*. (Quartiles are defined for sample sizes not of the form $4k+3$ as combinations of adjoining order statistics around 1/4 and 3/4 of the way up the ordering.) The *interquartile range* is defined by $IR = X_{(3k+3)} - X_{(k+1)}$. What is its breakdown point? Does it depend on the sample?

2. Suppose we want to estimate a scale parameter σ with $0 < \sigma < \infty$ (not necessarily a standard deviation). Each compact set in the parameter space is included in a set $1/M \leq \sigma \leq M$.

(a) For what samples $X_1, ..., X_n$ of real numbers is the median absolute deviation MAD in the parameter space?

(b) If the observations X_j are all different, what is the breakdown point of MAD with the given parameter space? *Hint*: if the parameter space is $[0, \infty)$, so that 0 is a possible value, the breakdown point is found in the proof of Theorem 3.44.5.

(c) Suppose that r of the X_j are the same for some $r = 2, ..., n$ and the rest are all different. Then what is the breakdown point of MAD at such a sample?

3. For the wide-sense Huber functions ρ on p. 6 of Section 3.4, (a), (b) and (c), and $\psi(x) = \rho'(x)$ for $x \neq 0$, $\psi(0) = 0$, find the number $A = \lim_{x \to +\infty} \psi(x)$ in each case.

4. Show that for $n = 2$, the M-estimate for location defined by any ψ function satisfying the hypotheses of Theorem 3.44.5 is $(X_1 + X_2)/2$.

5. Show that for the ρ function $\rho(x) = \sqrt{1 + x^2}$ and corresponding ψ function $\psi(x) = \rho'(x)$, we have $\theta^* = \theta^{**}$ for any sample.

REFERENCES

Frank R. Hampel, Peter J. Rousseeuw, Elvezio M. Ronchetti, and Werner A. Stahel (1986). *Robust Statistics: The Approach based on Influence Functions.* Wiley, New York.

Peter J. Huber (1981) *Robust Statistics.* Wiley, New York.