

Breakdown points of some 1-dimensional location estimators

Recall that a set C in a real vector space is called *convex* if for any x, y in C and $0 \leq \lambda \leq 1$ we have $\lambda x + (1 - \lambda)y \in C$. In the real line, a convex set is just an interval, a half-line or the whole line. At a finite endpoint it may be open or closed. A real-valued function f on the convex set C is called *convex* if for each x, y , and λ as before, we have $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$. Here f is called *strictly convex* if whenever $x \neq y$ and $0 < \lambda < 1$, “ \leq ” is replaced by “ $<$ ” in the definition of convex function.

For example, the function $f(x) = x^2$ on \mathbb{R} is strictly convex. The function $f(x) = |x|$ is convex, but not strictly convex.

A non-constant function $\rho(\theta, x)$ for x and θ real will be called a *wide-sense Huber function* if $\rho(\theta, x) \equiv \rho(|x - \theta|)$ where $\rho(x) \equiv \rho(-x)$, ρ is convex, and $\rho(x)/|x|$ is bounded as $|x| \rightarrow \infty$. The convexity and symmetry properties imply that ρ attains its absolute minimum at 0 (and perhaps elsewhere). Examples of wide-sense Huber functions include

- (a) $\rho(x) := |x|$,
- (b) $\rho(x) := (c^2 + x^2)^{1/2} - c$ for any real $c > 0$, and
- (c) $\rho(x) := x^2$ for $|x| \leq b$ and $\rho(x) := c|x| - d$ for $|x| > b$ where $b > 0$ and the other constants are chosen to make ρ continuously differentiable. Then $cb - d = b^2$ and $2b = c$, so $d = b^2$ and for $|x| > b$, $\rho(x) = b(2|x| - b)$.

Since Huber especially studied functions defined by (c), they might be called “narrow-sense Huber functions.”

The functions in (b) and (c) are strictly convex in neighborhoods of 0, and the ones in (b) are strictly convex everywhere. Note that if ρ is convex, then the sum $\sum_{i=1}^n \rho(X_i - \theta)$ is convex in θ for any X_i . Also, if ρ is strictly convex for $|x| \leq b$, then the sum is strictly convex in a neighborhood of θ for any θ such that $|X_i - \theta| < b$ for some i . For n large and b not too small, the set of such θ will often include the set on which the sum takes its smallest values, so that the minimum will be unique. We will always have uniqueness if ρ is strictly convex everywhere, as $(c^2 + x^2)^{1/2}$ is.

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 $\psi(t) \rightarrow A$ as $t \rightarrow +\infty$ for some $A > 0$. Examples of such functions ψ include the derivatives $\rho'(x)$ of wide-sense Huber functions, 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, \dots, X_n) , let

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

This is finite since the sum is ≤ 0 for $\theta \geq X_{(n)}$ and also > 0 when $\theta \leq X_{(1)} - t$ for some t

such that $\psi(t) > 0$. Analogously, define

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

which is also finite since the sum is ≥ 0 for $\theta \leq X_{(1)}$ and < 0 for $\theta \geq X_{(n)} + t$. We have $\theta^* \leq \theta^{**}$ because of the monotonicity of ψ . In order to have a unique estimator, the (location) *M-estimator* for the sample (based on ψ) will be defined, as for the sample median, by

$$\hat{\theta} := \hat{\theta}(X_1, \dots, X_n) := \frac{1}{2}(\theta^* + \theta^{**})(X_1, \dots, X_n).$$

It will be shown that such estimators have the same (finite sample) breakdown points as the median, converging to $1/2$ as $n \rightarrow \infty$ 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 (a specific scale estimator will be given below). The resulting estimator will be called $\hat{\theta}_S$. 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.

For 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 $S = \text{MAD}/.6745$, 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. as in Randles and Wolfe, Sec. 7.4.

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

Theorem. 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 $t \rightarrow \infty$ 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 $i = 1, \dots, n$, we have $Y_{(j)} \leq X_{(n)}$, so $\theta^{**}(Y_1, \dots, Y_n) \leq X_{(n)} + \kappa$ 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 2 of the handout “Introduction to robustness: breakdown points.” 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

$$(1) \quad Y_{(j)} \leq X_{(n)}$$

and if M_Y is the median of Y_1, \dots, 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 $\text{MAD}_Y \leq X_{(n)} - X_{(1)}$ and

$$(2) \quad S_Y := S(Y_1, \dots, Y_n) \leq K_X := (X_{(n)} - X_{(1)})/0.6745.$$

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 $\sum_{i=1}^n \psi((Y_i - \theta)/S_Y) < 0$ 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 will hold if $\theta > X_{(n)} + \kappa K_X$ by (1) and (2). So

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

Symmetrically, we have

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

So $\hat{\theta}(Y_1, \dots, Y_n)$ 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, \dots, Y_n containing more than $n/2$ of the original observations X_1, \dots, 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. \square

REFERENCE

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