The Annals of Statistics
1984, Vol. 12, No. 1, 119–126

FINITE SAMPLE BREAKDOWN OF *M*- AND *P*-ESTIMATORS[1]

BY PETER J. HUBER

Harvard University

The finite sample breakdown properties of M-estimators, defined by $\sum\rho(x_i - T) = \min!$, and of the associated Pitman-type or P-estimators, defined by

$$T = \frac{\int \exp\{-\sum\rho(x_i - \theta)\}\theta\, d\theta}{\int \exp\{-\sum\rho(x_i - \theta)\}\, d\theta},$$

are investigated. If ρ is symmetric, and $\psi = \rho'$ is monotone and bounded, then the breakdown point of either estimator is $\varepsilon^* = \frac{1}{2}$. If ψ decreases to 0 for large x ("redescending estimators"), the same result remains true if ρ is unbounded. For bounded ρ, the P-estimator is undefined, and the breakdown point of the M-estimator typically is slightly less than $\frac{1}{2}$; it is calculated in explicit form.

1. Introduction. Since 1970 many new types of robust location estimators have been introduced. Somewhat surprisingly, the breakdown point (Hampel 1968, 1971) of hardly any of them is known; see Donoho and Huber (1982) for a recent discussion of this concept. The present paper begins to fill this gap in the literature by establishing the breakdown properties of redescending M-estimators (see Andrews et al., 1972) and of Pitman-type or P-estimators (Johns, 1979).

For the purposes of this paper, we define the value $T(X)$ of an M-estimator, based on the sample $X = (x_1, \cdots, x_n)$, by the property that it produces an absolute minimum of

$$\sum\rho(x_i - T), \tag{1.1}$$

rather than by the conventional definition of $T(X)$ as the solution of

$$\sum\psi(x_i - T) = 0 \tag{1.2}$$

nearest to the sample median (or the like), where $\psi = \rho'$ is the derivative of ρ.

If ψ is monotone and bounded, with $\psi(-\infty) = -\psi(\infty)$, then the breakdown point of T, that is the smallest fraction of bad sample values that may cause the estimator to take on arbitrarily large values, is $\varepsilon^* = \frac{1}{2}$. The same is true if we scale the observations by the MAD (median absolute deviation from the median) and determine T from

$$\sum\rho\left(\frac{x_i - T}{\text{MAD}}\right) = \min! \tag{1.3}$$

Received October 1982; revised July 1983.

[1] This work was facilitated in part by National Science Foundation Grant MCS-79-08685 and Office of Naval Research Contract N00014-79-C-0512.

AMS 1980 *subject classification.* Primary 62F35.

Key words and phrases. Breakdown point, robustness, M-estimators, P-estimators, redescending estimators.

or from

$$\sum \psi\left(\frac{x_i - T}{\text{MAD}}\right) = 0. \tag{1.4}$$

For details see Huber (1981).

If ψ is nonmonotone and redescends to 0, we shall now show that the breakdown point remains $\varepsilon^* = \frac{1}{2}$ if ρ is unbounded. If ρ is bounded, the breakdown point can be less than $\frac{1}{2}$, but for reasonably well-tuned estimates the deficiency is negligibly small.

The associated P-estimates, defined by

$$T(X) = \frac{\int \exp\{-\sum \rho(x_i - \theta)\}\theta \, d\theta}{\int \exp\{-\sum \rho(x_i - \theta)\} \, d\theta}. \tag{1.5}$$

exist only for unbounded ρ; they have the same breakdown point $\frac{1}{2}$.

Since it appears that the notion of breakdown point is most useful in a small sample setup (cf. the discussion in Donoho and Huber [1982], and in Donoho [1982]), we shall give it a finite sample definition.

2. Finite sample breakdown. Let $X = (x_1, \cdots, x_n)$ be a finite sample of size n. We can corrupt this sample in many ways; in this paper we shall work exclusively with *ε-contamination*: adjoin m arbitrary values $Y = (y_1, \cdots, y_m)$ to the sample. The corrupted sample $X' = X \cup Y$ then has size $n + m$ and contains a fraction $\varepsilon = m/(n + m)$ of bad values.

We could also have corrupted the sample by *ε-replacement*: replace an arbitrary subset of size m of the sample by arbitrary values $y_1, \cdots, y_m$. The corrupted sample X' then has size n and contains a fraction $\varepsilon = m/n$ of bad values.

In unstructured problems (like location/scale estimation), ε-contamination usually is more convenient to deal with. In structured cases (e.g. for time series problems) the situation is reversed.

For a fixed type of ε-corruption we define the *maximum bias*

$$b(\varepsilon; X, T) = \sup |T(X') - T(X)|,$$

where the supremum is taken over all ε-corrupted samples X'.

The *breakdown point* ε^* is defined as

$$\varepsilon^*(X, T) = \inf\{\varepsilon \mid b(\varepsilon; X, T) = \infty\}.$$

The breakdown point can be as high as 1 (for a constant statistic, or more generally, for a Bayes estimate whose prior has compact support), and it can approach 0 (e.g. for the sample mean, $\varepsilon^* = 1/(n + 1)$).

The sample median has $\varepsilon^* = \frac{1}{2}$, which is the highest value attainable by a translation equivariant estimate (if $\varepsilon = \frac{1}{2}$, no translation equivariant estimate can decide whether X or Y is the good part of the sample, and thus it must break down.

3. Redescending M-estimates: bounded ρ. We first consider the fixed

scale case (1.1) and assume that ρ has a minimum at 0, $\rho(0) = -1$, that ρ increases monotonely toward both sides, and that $\lim \rho(x) = 0$ for $|x| \to \infty$.

THEOREM. *If we put*

$$\sum_X \rho(x_i - T(X)) = -A, \tag{3.1}$$

then the ε-contamination breakdown point of T is

$$\varepsilon^*(X, T) = \frac{m^*}{n + m^*}, \tag{3.2}$$

where m^* is an integer satisfying $\lceil A \rceil \leq m^* \leq \lfloor A \rfloor + 1$. If there is a $c < \infty$ such that $\rho(x) = 0$ for $|x| \geq c$, we have $m^* = \lceil A \rceil$.

PROOF. Assume first that the number of elements in Y is $m < A$; we shall show that then $T(X \cup Y)$ stays bounded. Let $\delta > 0$ be such that $m + n\delta < A$, and let c be such that $\rho(x) \geq -\delta$ for $|x| \geq c$. Let t be any real number such that $|x - t| \geq c$ for all x in X. Then

$$\sum_{x \in X \cup Y} \rho(x - T(X)) \leq -A \tag{3.3}$$

and

$$\sum_{x \in X \cup Y} \rho(x - t) \geq -n\delta - m. \tag{3.4}$$

Hence the left hand side of (3.3) is strictly smaller than the left hand side of (3.4), and it follows that $T(X \cup Y)$ must fall within distance c from a point in X.

On the other hand, if $m > A$, let $\delta > 0$ be such that $m - m\delta > A$, and let c be such that $\rho(x) \geq -\delta$ for $|x| \geq c$. Let y be any real number, and assume that all points in Y are equal to y. Then, for all t with $|y - t| \geq c$, we obtain

$$\sum_{x \in X \cup Y} \rho(x - t) \geq -A - m\delta, \tag{3.5}$$

and

$$\sum_{x \in X \cup Y} \rho(x - y) \leq -m. \tag{3.6}$$

Hence the left hand side of (3.6) is strictly less than the left hand side of (3.5), and it follows that $T(X \cup Y)$ must lie within a distance c from y. If we let $y \to \infty$, breakdown occurs.

This leaves the case $m = A$ open. But if $\rho(x) = 0$ for $|x| \geq c$, then (3.5) holds with $\delta = 0$, and if y is chosen sufficiently far out (namely such that $|x - y| \geq 2c$ for all $x \in X$), we easily verify that the minimum problem has the two solutions $T(X \cup Y) = T(X)$ and $T(X \cup Y) = y$. Thus, there is at least one solution going off to infinity, and we have breakdown. □

Clearly, the breakdown point as determined in the above theorem depends not only on the shape of ψ, and the scaling (tuning), but also on the sample configuration.

If scale is determined from sources extraneous to the sample—this is the typical situation in structured problems like regression and ANOVA—and if it is severely underestimated because of unrecognized heteroskedasticity, the breakdown point can become dangerously low.

However, for the usual redescending estimators and for reasonable tuning (i.e. such that the efficiency at the normal model is high and the gross error sensitivity is low, within reason) the breakdown point is quite high, usually above 0.4.

If scale is determined from the sample itself, the situation gets even more favorable with regard to breakdown. The presence of bad observations increases the MAD; while this increases the gross error sensitivity (the maximum of the sensitivity curve), it also increases the quantity A in the theorem, and numerical examples show that the breakdown point then typically exceeds 0.49, at least for the customary estimates with the customary tuning constants, e.g. for the biweight

$$\psi(x) = x\ (1 - (x/c)^2)_+^2$$

with $c = 6$.

4. Redescending M-estimates: unbounded ρ. Assume now that ρ is symmetric, $\rho(0) = 0$, and that ρ is increasing towards both sides. Assume

$$\lim_{|x|\to\infty} \rho(x) = \infty,$$

but

$$\lim_{|x|\to\infty} \frac{\rho(x)}{|x|} = 0.$$

Furthermore, we shall assume that $\psi = \rho'$ is continuous, and that there is an x_0 such that ψ is weakly increasing for $0 < x < x_0$, weakly decreasing for $x_0 < x < \infty$. (At the cost of some complications in the proofs, these regularity assumptions on ψ could be considerably weakened.)

THEOREM 4.1. *The ε-contamination breakdown point of an M-estimate* (1.1) *with ρ satisfying the above conditions is $\varepsilon^* = \frac{1}{2}$.*

We first prove two auxiliary lemmas. Let

$$M(t) = \sup_x |\rho(x + t) - \rho(x)|.$$

Since ρ is symmetric, we clearly have $M(-t) = M(t)$, and we may omit the absolute value bars in the definition without changing $M(t)$.

LEMMA 4.2. *The difference $\eta(t) = M(t) - \rho(t)$ is bounded: $0 \le \eta(t) \le x_0\psi(x_0)$. For $t \ge x_0$, we have $\eta(t) \le x_0\psi(t)$, hence $\eta(t) \to 0$ for $t \to \infty$.*

PROOF. $\eta(t) \ge 0$ is clear. Keep $t \ge 0$ fixed. Then $\rho(x + t) - \rho(x) = 0$ for $x = -t/2$, and the mean value theorem implies that $\rho(x + t) - \rho(x) = t\psi(x + \gamma t) \to$

0 for $x \to \infty$ (with $0 < \gamma < 1$). Hence, $\rho(x + t) - \rho(x)$ reaches its maximum at some x_1, $-t/2 < x_1 < \infty$; at x_1, its derivative vanishes, thus $\psi(x_1 + t) = \psi(x_1)$. In view of the representation

$$\rho(x + t) - \rho(x) = \int_x^{x+t} \psi(s)\, ds$$

and of the monotonicity properties of ψ, ψ must reach its maximum at a point between x_1 and $x_1 + t$. Without loss of generality, we may take x_0 to be this point. Then we must have

$$0 < x_1 < x_0 < x_1 + t,$$

and

$$\eta(t) = M(t) - \rho(t) = [\rho(x_1 + t) - \rho(t)] - [\rho(x_1) - \rho(0)]$$
$$\leq x_1 \psi(t + \gamma x_1) \quad \text{with} \quad 0 < \gamma < 1.$$

The assertion of the lemma follows.

LEMMA 4.3. *Let n and m be the respective sample sizes of X and Y. Put*

$$\Delta_{X \cup Y}(t) = \textstyle\sum_{x \in X \cup Y} \rho(x - t) - \rho(x).$$

Then there is a constant C which depends on X and on m, but not on the actual values in Y, such that for all t

$$(n - m)\rho(t) - C \leq \Delta_{X \cup Y}(t) \leq (n + m)\rho(t) + C.$$

PROOF. Write

$$\Delta_X(t) = \textstyle\sum_X [\rho(x - t) - \rho(x)] = n\rho(t) + \sum_X [\rho(x - t) - \rho(t)] - \sum_X \rho(x).$$

Since $|\rho(x - t) - \rho(t)| = |\rho(t) - \rho(t - x)| \leq |x|\psi(x_0)$, we have

$$|\Delta_X(t) - n\rho(t)| \leq C_1$$

with $C_1 = \sum_X \rho(x) + \sum_X |x| \psi(x_0)$.

On the other hand

$$|\Delta_Y(t)| = |\textstyle\sum_Y [\rho(x - t) - \rho(x)]| \leq mM(t) = m\rho(t) + m\eta(t).$$

Since $\eta(t)$ is bounded, it follows that

$$|\Delta_Y(t)| \leq m\rho(t) + C_2$$

for some C_2 and the assertion of the lemma holds with $C = C_1 + C_2$. □

PROOF OF THEOREM 4.1. Assume $m < n$. It follows from Lemma 4.3 that $\Delta_{X \cup Y}(t)$ is bounded away from 0 for sufficiently large t, uniformly in Y. Since $\Delta_{X \cup Y}(0) = 0$, and since $\Delta_{X \cup Y}$ reaches its absolute minimum at $T(X \cup Y)$, it follows that $T(X \cup Y)$ cannot be outside a certain bounded neighborhood of 0. Thus $\varepsilon^* > m/(n + m)$. It follows that $\varepsilon^* \geq \frac{1}{2}$. □

5. Pitman-type estimates: convex ρ. The Pitman-type, or P-, estimators of location are defined by

$$T_P = \frac{\int \{\Pi f(x_i - \theta)\}\theta \, d\theta}{\int \{\Pi f(x_i - \theta)\} \, d\theta} = \frac{\int \exp\{-\sum \rho(x_i - \theta)\}\theta \, d\theta}{\int \exp\{-\sum \rho(x_i - \theta)\} \, d\theta}, \tag{5.1}$$

where $\rho(x) = -\log f(x)$ is an essentially arbitrary function (f need not be a probability density). The only constraint is that the integrals in (5.1) should exist.

We say that the M-estimates (1.1) based on the same ρ is the M-estimate associated with a particular P-estimate.

Assume that ρ is symmetric and convex, and let $\psi = \rho'$ be the derivative of ρ. Assume that ψ is bounded.

LEMMA 5.1. *Under the above assumption T_P is monotone increasing in all of its arguments.*

PROOF. Denote the numerator and denominator of (5.1) by N and D respectively. It is easy to show by the dominated convergence theorem that we can differentiate under the integral signs, thus,

$$\frac{\partial T}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{N}{D}\right) = \frac{-D \int \psi(x_i - \theta)\exp\{\cdots\}\theta \, d\theta + N \int \psi(x_i - \theta)\exp\{\cdots\} \, d\theta}{D^2}$$

$$= \frac{\int (T_P - \theta)\psi(x_i - \theta)\exp\{\cdots\} \, d\theta}{D}.$$

Note that

$$\int (T_P - \theta)\exp\{\cdots\} \, d\theta = 0,$$

hence we may write

$$\frac{\partial T_P}{\partial x_i} = \frac{\int (T_P - \theta)[\psi(x_i - \theta) - \psi(x_i - T_P)]\exp\{\cdots\} \, d\theta}{D}.$$

Since ψ is monotone, the integrand in this expression is positive, hence $\partial T_P/\partial x_i \geq 0$. □

Note that we may replace $\rho(x_i - \theta)$ by $\rho(x_i - \theta) - \rho(x_i)$ in the definition of T_P. Furthermore, convexity of ρ implies that $\rho(x) - \rho(x - \theta)$ is an increasing function of x, and that

$$\lim_{x \to \infty} \rho(x - \theta) - \rho(x) = -c\,\theta,$$

where $c = \sup \psi(x) < \infty$.

On the other hand

$$\frac{\rho(x - \theta) - \rho(x)}{\theta} = -\frac{1}{\theta}\int_0^\theta \psi(x - t) \, dt,$$

and it follows

$$\lim_{\theta\to\pm\infty} \frac{\rho(x-\theta)-\rho(x)}{|\theta|} = c.$$

It follows from Lemma 5.1 that the maximum bias that can be caused by m contaminating observations occurs if they are all put at $+\infty$, and it follows from the preceding remarks that its value is

$$b\left(\frac{m}{n+m}, X, T_P\right) = T_P(X \cup Y) - T_P(X)$$

$$= \frac{\int \exp\{-\sum_X [\rho(x-\theta)-\rho(x)] + mc\theta\}\theta\, d\theta}{\int \exp\{-\sum_X [\rho(x-\theta)-\rho(x)] + mc\theta\}\, d\theta} - T_P(X).$$

Moreover, for large $|\theta|$ the exponent in this formula is

$$= \{-n|\theta|(1+o(1)) + m\theta\}c,$$

and therefore both integrals are finite so long as $m < n$. Hence the breakdown point is $\varepsilon^* \geq \frac{1}{2}$. Thus we have proved (assuming that ρ is differentiable):

THEOREM 5.2. *Assume ρ is symmetric and convex, and $0 < \lim_{|x|\to\infty}\rho(x)/|x| = c < \infty$. Then the breakdown point of the P-estimate* (5.1) *is $\varepsilon^* = \frac{1}{2}$.*

PROOF. If ρ is not differentiable, the proof needs straightforward but tedious modifications, which are left to the reader. □

6. Pitman-type estimates: non-convex ρ. Assume now that ρ is no longer convex, but satisfies the regularity conditions of Section 4. We define, as in Section 5,

$$T_P = \frac{\int \exp\{-\sum [\rho(x_i-\theta)-\rho(x_i)]\}\theta\, d\theta}{\int \exp\{-\sum [\rho(x_i-\theta)-\rho(x_i)]\}\, d\theta}. \tag{6.1}$$

Clearly, bounded functions ρ do not make sense here; in order that the P-estimate exists we must require that there is a constant $k > 0$ such that

$$\int \exp\{-k\rho(\theta)\}|\theta|\, d\theta < \infty.$$

In particular, we can take $k = 1$ if the P-estimate exists for sample size 1.

THEOREM 6.1. *The ε-contamination breakdown point of the P-estimate satisfies $\varepsilon^* > (n-k)/(2n-k)$; if $k \leq 1$, we have $\varepsilon^* = \frac{1}{2}$.*

PROOF. In the notation of Lemma 4.3, the exponent in (6.1) is $\Delta_{X\cup Y}(\theta)$, and it follows from this lemma that

$$|T_P(X \cup Y)| \leq \frac{\int \exp\{-(n-m)\rho(\theta) + C\}|\theta|\, d\theta}{\int \exp\{-(n+m)\rho(\theta) - C\}\, d\theta}.$$

If $n - m \geq k$, the right hand side is finite and provides a bound on $T(X \cup Y)$ which is uniform in Y. Hence T_P does not break down for $m \leq n - k$, and it follows that $\varepsilon^* > (n - k)/(2n - k)$. If $k \leq 1$, this implies that the smallest m for which breakdown can happen is $m = n$, hence $\varepsilon^* = \frac{1}{2}$.

Acknowledgment. I thank Helmut Rieder for pointing out several inaccuracies in the original manuscript.

REFERENCES

ANDREWS, D. F. *et al. (1972). Robust Estimates of Location.* Princeton Univ. Press, N.J.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. (Manuscript).

DONOHO, D. L. and HUBER, P. J. (1982). The notion of breakdown point. In: *Festschrift in Honor of Erich Lehmann,* Ed. by K. Doksum and J. L. Hodges, Wadsworth, Belmont, CA.

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

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

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

JOHNS, M. V. (1979). Robust Pitman-like estimators. In: *Robustness in Statistics,* Ed. by Launer and Wilkinson. Academic, New York.

DEPARTMENT OF STATISTICS
SCIENCE CENTER
HARVARD UNIVERSITY
ONE OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138